A Remote Sensing-Based Tool for Assessing Rainfall-Driven Hazards

Daniel B. Wright[1*]
Ricardo Mantilla[2]
Christa D. Peters-Lidard[3]

[1]University of Wisconsin-Madison, Madison, Wisconsin
[2]University of Iowa, Iowa City, Iowa
[3]NASA Goddard Space Flight Center, Greenbelt, Maryland

[*]danielb.wright@wisc.edu

Abstract

RainyDay is a Python-based platform that couples rainfall remote sensing data with Stochastic Storm Transposition (SST) for modeling rainfall-driven hazards such as floods and landslides. SST effectively lengthens the extreme rainfall record through temporal resampling and spatial transposition of observed storms from the surrounding region to create many extreme rainfall scenarios. Intensity-Duration-Frequency (IDF) curves are often used for hazard modeling but require long records to describe the distribution of rainfall depth and duration and do not provide information regarding rainfall space-time structure, limiting their usefulness to small scales. In contrast, RainyDay can be used for many hazard applications with 1-2 decades of data, and output rainfall scenarios incorporate detailed space-time structure from remote sensing. Thanks to global satellite coverage, RainyDay can be used in inaccessible areas and developing countries lacking ground measurements, though results are impacted by remote sensing errors. RainyDay can be useful for hazard modeling under nonstationary conditions.

Keywords: scenarios; extreme rainfall; remote sensing; floods; landslides; risk assessment

Software Availability

Name of Software: RainyDay Rainfall Hazard Modeling System

Developer: Daniel B. Wright

Contact: Daniel B. Wright; Address: Room 1269 Engineering Hall, 1415 Engineering Drive, Madison, WI 53706, USA; Email: danielb.wright@wisc.edu

Year first available: 2015

Required hardware and software: RainyDay requires Python 2.7 or newer (not tested with Python 3.0 or higher) with Numpy and Scipy. The Netcdf4 and GDAL APIs and Python libraries are also required. RainyDay will run on Macintosh, Linux, and Windows machines with the proper APIs and Python libraries.

Cost: Free. RainyDay is currently available by request at https://bitbucket.org/danielbwright/rainyday. Open-source release under version 3.0 of the GNU General Public License (http://www.gnu.org/licenses/gpl-3.0.en.html) is planned, with unrestricted public access to the code repository.

1. Introduction

Rainfall-driven hazards such as floods and landslides are the most common natural disasters worldwide, and amongst the most devastating. A growing number of computational hazard models are available to transform extreme rainfall inputs into hazard predictions, including distributed hydrologic models for the movement of water into and through river systems (e.g., Smith et al., 2004); hillslope stability and run-out models for landslide initiation and subsequent motion (e.g. Brenning, 2005 and Preisig and Zimmermann, 2010, respectively); and hydraulic models for flood wave propagation in channels and floodplains (e.g., Horritt and Bates, 2002). These models have seen significant advances in recent decades, and have become key components in probabilistic hazard and risk assessment in fields such as natural catastrophe risk insurance, infrastructure design, and land-use planning. The hazard predictions produced by these models tend to be highly sensitive to the amount, timing, and spatial distribution of rainfall inputs. Unfortunately, progress on developing realistic rainfall inputs for probabilistic hazard and risk assessment has been relatively limited. This paper introduces RainyDay, a Python-based platform that addresses this shortcoming by coupling rainfall remote sensing data from satellites or other sources with a technique for temporal resampling and spatial transposition known as Stochastic Storm Transposition (SST) to generate highly realistic probabilistic rainfall scenarios.

Rainfall inputs for long-term hazard and risk assessment require a probabilistic description of three interrelated components: duration, intensity, and space-time structure. Efforts to jointly model these components are usually referred to as *rainfall frequency analysis*, a simple term that belies the complexity of the physical phenomena and analytical methods involved. The probability structure of the first two components, rainfall duration and intensity, has been a focus of research and application for decades (see U.S. Weather Bureau, 1958 and Yarnell, 1935 for early examples). These two components are strongly linked and together they determine the probability distribution of rainfall volume (or depth) at a point or over an area. The third component, space-time structure, describes the spatial and temporal variability of rainfall and is determined by storm size, horizontal velocity, and the temporal evolution of spatial rainfall coverage. Space-time structure can thus be understood as describing the “when” and “where” of extreme rainfall, whereas intensity and duration describe “how much.”

Rainfall space-time structure can be an important hazard determinant. For example, a rainstorm that is short-lived and small in spatial extent may pose a significant flash flood threat in a narrow mountain valley or urban area, but may not represent a hazard on a larger river system. Conversely, a month-long rainy period could lead to flooding on a major river due to the gradual accumulation of water in soils, river channels, and reservoirs, but may never feature a short-lived "burst" of rainfall sufficiently intense to cause flash flooding at smaller scales. Similarly, a storm that covers a large area or passes over a series of valleys could lead to more widespread landslide or debris flow occurrences than a smaller or stationary storm. Rainfall space-time structure and its importance as a hazard trigger, therefore, must be understood within the context of the particular geography and scale in question. Due to its complexity, rainfall space-time structure has traditionally been less well understood than intensity and duration, and its representation in hazard modeling has been less sophisticated.

The probability distribution of rainfall volume for a given duration is usually derived from rain gages and distilled into Intensity-Duration-Frequency (IDF) curves, such as those provided by the National Oceanic and Atmospheric Administration's (NOAA) Atlas 14 (Bonnin et al., 2004). Long records (spanning many decades) are generally needed to define the extreme tail of such distributions. The challenge of measuring extreme rainfall over long time periods and over large areas using rain gages has hindered IDF estimation in many developed countries, while the lack of data in poor countries and in inaccessible terrain means that IDF estimation using such methods is virtually impossible in many locations. Furthermore, the ability to measure rainfall space-time structure at a high level of detail using dense networks of rain gages is nonexistent outside of a handful of wealthy cities and research-oriented observation networks. "Regionalization,"—the pooling of hazard information over a larger area in order to inform analyses at particular locations (see, e.g. Alexander, 1963 for an early discussion of rainfall regionalization and Stedinger et al., 1993 for a review)—has helped with IDF estimation using short records in areas where rain gage densities are moderate or high. These techniques offer little help, however, in parts of the world where rain gages are few or nonexistent, and do not offer a framework for incorporating rainfall space-time properties into hazard estimation. Even where long rainfall records do exist, nonstationarity due to climate change may mean that earlier

110 portions of the record are no longer representative of current or future IDF properties (e.g. Cheng
111 and AghaKouchak, 2014).
112
113 Several techniques, which generally fall under the term of *design storm methods*, are used in
114 long-term hazard estimation to link IDF properties to space-time structure for probabilistic flood
115 hazard assessment (commonly referred to as *flood frequency analysis*). One central design storm
116 concept is to link rainfall duration to rainfall intensity via a measure of flood response time, such
117 as the time of concentration (e.g. McCuen, 1998). Another attempts to estimate area-averaged
118 rainfall from point-scale rainfall estimates using area reduction factors (ARFs; U.S. Weather
119 Bureau, 1958). Yet another uses dimensionless time distributions such as the family of U.S. Soil
120 Conservation Service 24-hour rainfall distributions (e.g. McCuen, 1998). Each of these methods
121 is highly empirical, laden with assumptions (see Wright et al., 2014a; Wright et al., 2014b;
122 Wright et al., 2013), valid only in certain contexts, and often misunderstood or misused (K.
123 Potter, personal communication, May 6, 2015).
124
125 SST explicitly links IDF properties rainfall space-time properties, providing certain advantages
126 over design storm methods. Similar to other regionalization techniques, SST aims to effectively
127 "lengthen" the period of record by using nearby observations, albeit using a fundamentally
128 different approach involving temporal resampling and spatial transposition of rainstorms drawn
129 from a catalog of observed rainfall events from the surrounding region. The inclusion of nearby
130 storms at least partially addresses the difficulty of accurately estimating rainfall hazards using
131 short records. SST can be used to estimate rainfall IDF properties and also to facilitate modeling
132 of interactions of rainfall space-time structure with geographic features (such as hillslopes and
133 river networks) at the appropriate spatial and temporal scales. It accomplishes this by generating
134 large numbers of extreme rainfall "scenarios," each of which has realistic rainfall structure based
135 directly on observations.
136
137 Alexander (1963), Foufoula-Georgiou (1989), and Fontaine and Potter (1989) describe the
138 general SST framework, while Wilson and Foufoula-Georgiou (1990) use the method for rainfall
139 frequency analysis and Gupta (1972) and Franchini et al. (1996) use it for flood frequency
140 analysis. In those days, however, the method was of limited practical use due to the lack of

detailed rainfall datasets with large areal coverage. Those studies also did not focus on the aspects of SST related to rainfall space-time structure nor its implications for hazard modeling.

The recent advent of satellite-based remote sensing provides a relatively low-cost means of measuring extreme rainfall over large parts of the globe at moderately high spatial and temporal resolution (30 minutes to 3 hours, 4 km to 25 km), while ground-based weather radar offers higher-resolution estimates (5-60 minutes, typically 1 to 4 km) over smaller regions. While the accuracy of rainfall remote sensing can be poor (particularly in cases of satellite-based estimates, e.g. Mehran and AghaKouchak, 2014; and mountainous regions, e.g. Nikolopoulos et al., 2013, Stampoulis et al., 2013), such data nonetheless offer unprecedented depictions of rainfall over large areas. This creates a variety of opportunities for hazards research and practice at various scales, ranging from forecasting and post-event analysis to long-term hazard assessment.

In the context of SST, the ongoing accumulation of remote sensing data to lengths of 10-20 years or more "unlocks" many of the as-yet unrealized opportunities offered by SST. Wright et al. (2013) demonstrated the coupling of SST with a 10-year high resolution radar rainfall dataset for IDF estimation, and the method was extended to flood frequency analysis for a small urban watershed using a distributed hydrologic model in Wright et al. (2014b). These two papers, along with Wright et al. (2014a) show that commonly-used design storm practices (ARFs, dimensionless time distributions) have serious shortcomings in representing the multi-scale space-time structure of extreme rainfall and critical interactions with of this structure with watershed and river network features. Wright et al. (2014b) also show that when SST is coupled with rainfall remote sensing data and a distributed hydrologic model, it can reproduce the role that this structure plays in determining multi-scale flood response. The RainyDay software described in this paper was developed to facilitate the use of SST in conjunction with rainfall remote sensing data.

Though SST was developed in the context of flood hazard estimation, it may prove useful for rainfall-triggered landslides and other mass movements, subject to the limited accuracy of remote sensing data in steep terrain and other limitations that will be discussed subsequently. Rainfall space-time structure governs the temporal distribution of rainfall volume onto individual

hillslopes, as well as the number of hillslopes subject to rainfall. In addition, steep landslide-prone terrain often has poorer rain gage coverage than lowland areas due to limited accessibility, suggesting that remote sensing rainfall estimates are potentially useful in such regions, particularly if improvements in accuracy can be realized (e.g. Shige et al., 2013).

Section 2 provides a description of the SST methodology. Section 3 discusses the specific implementation of SST in RainyDay, and some of the software's important features. Section 4 provides sample results from RainyDay and sensitivity analyses using different input rainfall datasets for rainfall and flood frequency analysis in order to illustrate its capabilities and some of its limitations, including for flood frequency analysis in nonstationary conditions. Section 5 includes discussion and concluding remarks.

2. The SST Methodology

In this section, we provide a step-by-step methodology for SST-based rainfall frequency analysis for a user-defined geographic "area of interest," *A* of arbitrary shape. Higher-level description of software features is left to Section 3, but it merits mention that in RainyDay, *A* can be a single remote sensing pixel, a rectangular area containing multiple pixels, or a contiguous area defined by a user-supplied polygon shapefile.

The following five steps describe the SST methodology, as implemented in RainyDay:

1. Identify a geographic transposition domain *A'* that encompasses the area of interest *A*. One could confine *A'* to regions with homogeneous extreme rainfall properties, (e.g. flat areas far from large water bodies and topographic features). However, such homogeneity would likely be difficult to rigorously determine in practice and regardless, such strict interpretation is likely to be overly limiting. RainyDay offers several diagnostic aids, discussed in Section 3.3, that help the user to understand rainfall heterogeneity over the region *A'* and to improve the performance of the SST procedure in cases where rainfall heterogeneities do exist. Additional issues related to the selection of *A'* are explored in Section 4.3.

2. Identify the largest m temporally non-overlapping storms in A' from an n-year rainfall remote sensing dataset, in terms of rainfall accumulation of duration t and with the same size, shape, and orientation of A. For example, the principal axis of the Turkey River watershed in northeastern Iowa in the central United States is oriented roughly northwest-southeast and has an area of 4400 km^2. In this case, the m storms are those associated with the m highest t-hour rainfall accumulations over an area of 4400 km^2 with the same size, shape, and orientation as the Turkey River watershed. We refer to this set of storms henceforth as a "storm catalog," with the same geographic extent as A' and the same spatial and temporal resolution as the input rainfall data. We refer to the m storms in the storm catalog henceforth as "parent storms." In RainyDay, the user can specify whether to exclude certain months (such as wintertime) from the storm catalog. Previous studies have shown that there can be low bias introduced in high-exceedance probability (i.e. frequent, low-intensity) events if m is small (e.g. Foufoula-Georgiou, 1989; Franchini et al., 1996; Wilson and Foufoula-Georgiou, 1990; see Wright et al., 2013 for a discussion). The sensitivity of SST results to the choice of m and A' is explored in detail in Section 4.3, but $m \approx 10n$ generally minimizes the low bias for frequent events, and would likely be a good starting point for new analyses. Low exceedance probability (i.e. rare) events are less sensitive to the choice of m (see Section 4.3).

   In RainyDay, duration t is a user-defined input, and as long as t is neither very short nor very long relative to the time scales of hazard response in A, subsequent hazard modeling results will be relatively insensitive to the chosen value. In this respect, the duration t in SST differs conceptually from design storm methods, in which hazard response is intrinsically sensitive to the user-specified duration, and this feature is indeed one of the chief advantages of SST over design storm methods for multi-scale flood hazard estimation (see Wright et al., 2014b for analysis and discussion). In the case of SST-based flood frequency analysis, t should be at least as long as the time of concentration and preferably somewhat longer.

3. Randomly generate an integer k, which represents a "number of storms per year." In previous SST literature, the assumption was made that k follows a Poisson distribution

with a rate parameter λ storms per year. The *m* parent storms are selected such that an average of λ = *m*/*n* storms per year are included in the storm catalog. For example, if *m* = 100 storms selected from a ten-year remote sensing dataset, then λ = 100/10 = 10.0 storms per year. RainyDay will generate *k* using either the Poisson distribution or an empirical distribution, discussed in Section 3.3. If the Poisson distribution is selected, RainyDay will automatically calculate λ based on user-specified *m* and the length of the input dataset.

4. Randomly select *k* parent storms from the storm catalog. For each selected parent storm, transpose all rainfall fields associated with that storm by an east-west distance Δ*x* and a north-south distance Δ*y,* where Δ*x* and Δ*y* are drawn from the distributions $D_X(x)$ and $D_Y(y)$ which are bounded by the east-west and north-south extents of *A'*, respectively. The motion and structure of the parent storm is unaltered during transposition and only the location is changed. The distributions $D_X(x)$ and $D_Y(y)$ were taken to be uniform in Wright et al. (2013) and Wright et al. (2014b), but RainyDay offers additional options, described in Section 3.3. We illustrate this step schematically in Figure 1. For each of the *k* transposed storms, compute the resulting *t*-hour rainfall accumulation averaged over *A*.

   Step 4 can be understood as temporal resampling and spatial transposition of observed storm events within a probabilistic framework to synthesize one year of heavy rainfall events over *A'* and, by extension, over *A*. RainyDay and previous SST efforts retain the largest (in terms of rainfall intensity) of the *k* events for subsequent steps and discard the *k-1* remaining events, though in principle these events could be retained. The single retained storm can be understood as a "synthetic" annual rainfall maximum, analogous to those annual rainfall maxima that are extracted from rain gage records for rainfall frequency analysis. It should be noted that these rainfall events do not form a continuous series, meaning that neither inter-storm periods nor the sequencing of the *k* storms are considered.

5. Repeat steps 3 and 4 a user-specified T_{max}, number of times, in order to create T_{max} years of *t*-hour synthetic annual rainfall maxima for *A*. RainyDay then assigns each annual

maximum a rank i according to its rainfall intensity relative to all others. Each of these ranked maxima can then be assigned an annual exceedance probability p_e^i where $p_e^i \equiv i/T_{max}$. Exceedance probability p_e is the probability in a given year that an event of equal or greater intensity will occur. The "return period" or "recurrence interval" T_i, commonly used in hazard analysis, is simply $T_i \equiv 1/p_e^i$, so if $T_{max} = 10^3$, it is possible to directly infer exceedance probabilities of $1.0 \geq p_e \geq 10^{-3}$ (recurrence intervals of $1 \leq T_i \leq 10^3$). Each of these rainfall events can then serve as one datum of an empirical IDF estimate or as a rainfall scenario for hazard modeling.

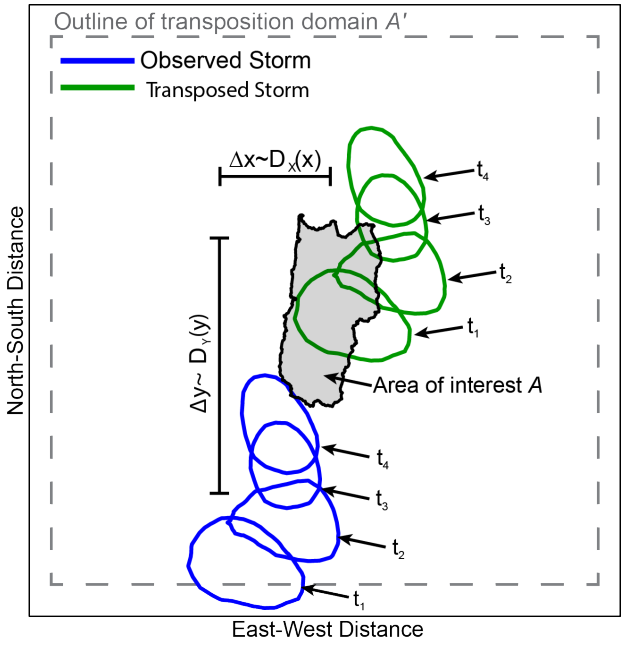


Figure 1: Depiction of SST procedure for a single storm consisting of four time intervals $t_1 \ldots t_4$. The blue ellipses illustrate the time evolution of an arbitrary rainfall isohyet derived from remote sensing observations, while the green ellipses show the time evolution of this same isohyet after transposition. Adapted from Wright et al. (2013).

3. RainyDay Software

3.1 Overview of Software

We wrote RainyDay to render SST more accessible and to streamline the code for speed and ease-of-use using Python. The majority of subroutines utilize the Scipy (Jones et al., 2011) and Numpy packages (Walt et al., 2011). To enhance speed, certain RainyDay subroutines call C

283 code through Scipy's "weave" functionality
284 (http://docs.scipy.org/doc/scipy/reference/tutorial/weave.html). Figure 2 shows a schematic of
285 workflow in RainyDay.
286
287 While the ranking of rainfall events described in Step 5 of the SST methodology in Section 2 is
288 based on rainfall intensity averaged over A, RainyDay will create NetCDF4 files
289 (http://www.unidata.ucar.edu/software/netcdf) that contain the transposed rainfall scenarios with
290 full depictions of rainfall space-time structure at the native spatial and temporal resolution of the
291 input. This is an important feature because space-time structure, and not just average rainfall
292 intensity over area A and duration t, is important in determining hazard response. For example,
293 one rainfall scenario may produce a more severe flood response than another scenario, even if it
294 has a lower overall average rainfall intensity over A and t, due to interactions with watershed
295 features (see Section 3.2 of this paper for discussion and Wright et al., 2014b for analysis).
296
297 We will provide the RainyDay source code, examples, and user documentation upon request, and
298 intend to release it under version 3 of the GNU General Public License
299 (http://www.gnu.org/copyleft/gpl.html) once we have completed sufficient testing and
300 documentation. The code is currently not parallelized, but shared-memory parallelization may
301 be added in the future. Computational time is determined mainly by the size of the input dataset
302 (record length n, input resolution, and geographic size of A and A'), while other factors, such as
303 m, t, T_{max}, and N can impact runtime. Computational speed, even without parallelization, is not
304 prohibitive on a modern desktop or laptop computer (several seconds to several hours for typical
305 configurations and input datasets).
306
307 To ensure accessibility for users inexperienced with Python, all of the necessary Python modules
308 are supported within recent versions of the Anaconda Python distribution from Continuum
309 Analytics (https://store.continuum.io/cshop/anaconda). The user must install NetCDF4 libraries
310 and any requisite dependencies. If the user wishes to use shapefile functionality, necessary for
311 defining A to be a shape other than a rectangle or a single rainfall pixel, the GDAL library
312 (http://www.gdal.org) and any necessary dependencies must also be installed.

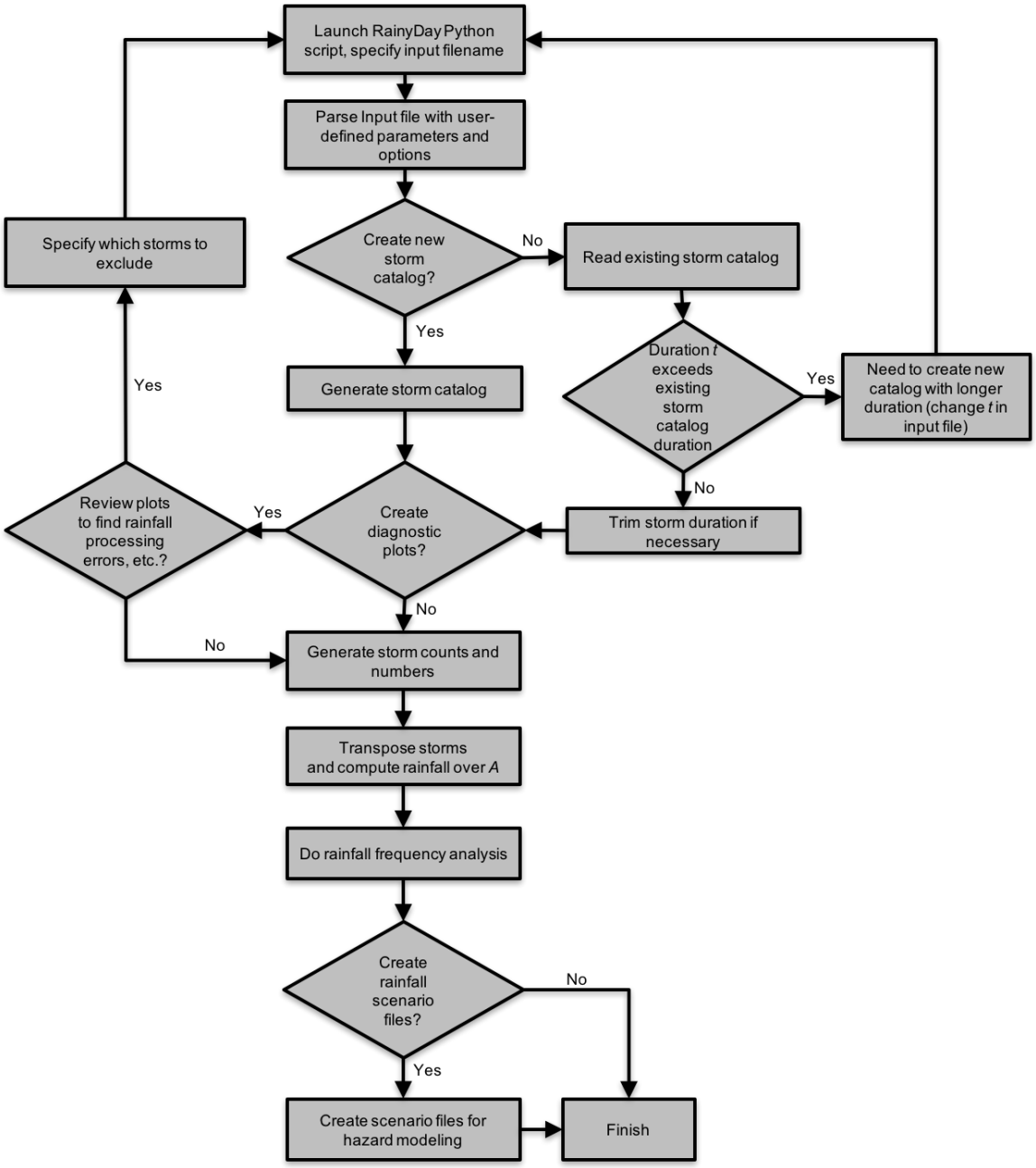


Figure 2: Flow chart demonstrating the workflow of RainyDay.

3.2 SST Internal Variability

In RainyDay, the user specifies N, the number of T_{max}-year long "ensemble members" to be generated. This enables the examination of "internal variability," i.e. how much variation in rainfall intensity is possible for a given p_e for a given input rainfall dataset and set of user-defined parameters. For example, if the user specifies $T_{max} = 10^3$ and $N = 100$, then there will be

100 intensity estimates for each p_e between 1.0 and 10^{-3}. RainyDay will automatically generate a text file containing the results of this rainfall frequency analysis, including the rainfall mean, minimum, and maximum (or, optionally, a quantile interval) for each p_e, computed from the N ensemble members.

If the scenarios generated by RainyDay are fed through a hazard model, then the ensemble spread will propagate through to generate ensemble hazard estimates. A useful and interesting feature of SST and RainyDay that is not examined in this paper, but is discussed at length in Wright et al. (2014b), is that the exceedance probability of rainfall and of subsequent hazards can be decoupled using SST, particularly if some realistic scheme is used to account for the initial conditions in A (such as soil moisture or baseflow). Consider the example where $N=1$ and 10^3 rainfall scenarios ($T_{max}=10^3$) are created as input to a distributed flood hydrologic model. One of these rainfall scenarios has $p_e=0.01$ (in terms of watershed-average t-hour rainfall depth over an area A). Even if initial conditions are kept constant across all T_{max} simulations, the p_e of the peak discharge or volume predicted by the model for this particular scenario need not be equal to 0.01, since the space-time structure of the rainfall scenario and its interactions with watershed and river network features can dampen or magnify the flood severity. If variability in initial conditions within the hazard model are considered, this dampening or magnification effect can be even greater. This property of SST contrasts with design storm methods, which typically assume a 1:1 relationship between the p_e of rainfall and the resulting hazard, though variability in initial conditions could in principle be used with design storm approaches to produce some degree of "decoupling" of rainfall and hazard p_e. Setting $N \geq 2$ allows for examination of differences in hazard p_e for a given rainfall p_e, or vice versa, which could lead the way to more detailed examination of the role of rainfall space-time structure (see Wright et al. 2014b) or initial conditions in probabilistic hazard estimation. RainyDay provides one simple scheme for creating variability in initial conditions, described in Section 3.5.

It should be pointed out that the ensemble spread generated in RainyDay is not completely comparable to the confidence intervals of more traditional rainfall or flood frequency analyses. The latter show statistical uncertainty associated with parameter estimation, which can be derived in different ways (e.g. bootstrapping, profile likelihood, etc.). Therefore, it might not be

reasonable to expect that the uncertainty ranges produced by RainyDay to be comparable to the confidence intervals of other IDF estimates. Like nearly all frequency analyses and IDF estimation methods, the ensemble spread generated by RainyDay does not consider measurement error, which, as mentioned previously, can be substantial. Since the ensemble spread is for a given set of user-defined values such as *A'* or *m*, it does not consider uncertainty associated with these choices. Analyses in Section 4.3 show how such uncertainties can be assessed, but fundamentally this requires manipulating the size or composition of the storm catalog through the choice of user-defined values, necessitating multiple distinct runs of RainyDay.

Ensemble spread is shown throughout Section 4 to illustrate various aspects of SST-based rainfall and flood frequency analysis. If the user is only interested in examining internal variability of SST-based rainfall IDF, then the number of ensemble members can be large (e.g. $N \geq 100$). If the user wishes to perform hazard simulations, however, *N* should be selected with consideration of the computational cost associated with large numbers of simulations, which can be substantial depending on the particular hazard model. To help manage the number of simulations required, the user can specify a rainfall return period threshold, below which output scenarios will not be created. For example, if the user specifies a 5-year threshold, no rainfall scenarios with a rainfall depth less than the 5-year return period depth will be written, which reduces the number of hazard simulations by 80% for a given value of *N* while still retaining the most extreme scenarios.

3.3 Rainfall Heterogeneity and Non-Uniform Spatial Transposition

A common criticism of SST is that its validity is restricted to regions with homogenous extreme rainfall properties. As previously mentioned, depending on how rigidly this criterion is enforced, the method would be limited to small, flat regions far from topographic features, water bodies, etc. It is unclear how homogeneity would be determined, particularly with the paucity of extreme rainfall data in most regions. Instead, steps can be taken to use SST in more varied geophysical settings. Regardless of the setting, the selection of *A'* requires an understanding of regional rainfall patterns and of the intrinsic assumptions of SST. Though more work is needed to understand the geographic limits of the applicability of RainyDay in complex terrain, the work of England et al. (2014) provides an example of SST in complex terrain.

RainyDay provides several tools to help understand the issue of rainfall heterogeneity, and, to some extent, to mitigate it. First, RainyDay produces a map showing the location of the rainfall centroids for all storms in the storm catalog, overlaid on a smoothed field of the spatial probability of storm occurrence within *A'*. This spatial probability of occurrence map is generated by applying a two-dimensional Gaussian kernel smoother to the (*x,y*) locations of the rainfall centroids for all the storms in the storm catalog. This smoothed field is then normalized such that the sum of all grid cells is 1.0, thus creating a two-dimensional probability density function of storm occurrence. A second plot shows these rainfall centroids overlaid with the average rainfall per storm across *A'*. These diagnostic plots assist in understanding regional variations in storm occurrences and rainfall over *A'*. Examples of these diagnostic plots for a region *A'* encompassing most of the state of Iowa in the central United States are shown in Figure 3. The top panel suggests that storms are somewhat more frequent in the southernmost third or so of the transposition domain (top panel), along with slightly elevated activity in the northeast quadrant. The bottom panel shows somewhat higher average storm rainfall in these two areas. Caution should be taken when drawing firm conclusions from these diagnostic plots, however, since rainfall heterogeneities evident in both storm occurrences and average storm rainfall may be the result of spatial biases in rainfall remote sensing estimates or of randomness in the climate system over the relatively short remote sensing record, rather than from "true" heterogeneity in the underlying rainfall hydroclimate.

Additional optional diagnostic outputs include static and animated rainfall maps for each storm in the storm catalog (not shown). These storm rainfall maps are useful for diagnosing "bad data," particularly in rainfall datasets that use ground-based weather radar contaminated by radar beam blockage and other unrealistic artifacts. RainyDay allows for the exclusion of user-identified storm periods from subsequent analysis, though anomalous periods must be identified y the user (i.e. no automatic data quality checking is provided).

The two-dimensional density function of spatial storm probability of storm occurrence can optionally be used as the basis for non-uniform spatial transposition (providing the $D_X(x)$ and $D_Y(y)$ described in Step 4 and Figure 1 in Section 2) so that the spatial distribution of storm

occurrences will be preserved between the input data and output rainfall scenarios and IDF estimates. Section 4.3 examines the impact of this optional feature on results for the Iowa study region, along with potential implications.

It is important to note that this approach only addresses the spatial heterogeneity of storm occurrences, not of spatial variations in the climatology of rainfall intensity (due to topography or other factors). For example, if *A'* contains both a flat plain and an adjacent mountain range, the probability of storm occurrence will vary across *A'*. This variation will be captured in the two-dimensional density function of spatial storm probability and, using the optional non-uniform spatial transposition scheme, will be reflected in RainyDay outputs. In this example, rainfall intensity from these storms will also vary according to the underlying topography. The current transposition scheme in RainyDay cannot explicitly account for this intensity variation. This is likely to be a serious constraint in many regions.

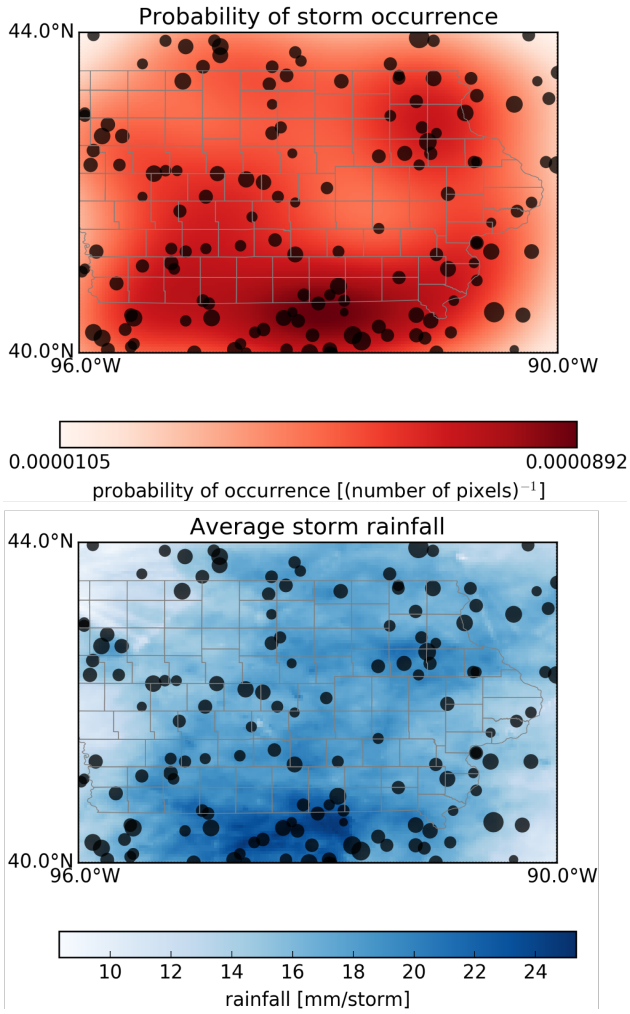


Figure 3: Example of diagnostic plots produced by RainyDay for 24-hour duration rainfall from the Stage IV rainfall dataset (described in Section 4.1) over a region encompassing the state of Iowa, United States. Top: shading indicates spatial probability of storm occurrence. Bottom: shading indicates the average rainfall per storm from the same storm catalog. Black dots show the rainfall centroids for each storm in the storm catalog. Dot size in both panels indicates relative rainfall storm total rainfall depth. Key RainyDay parameters: m=150 storms, A'=[40° to 44° N, 90° to 96° W]. A is a single Stage IV rainfall pixel (approximately 16 km^2), T_{max}=1000, and t=24 hours.

3.4 Empirical Temporal Resampling

As mentioned in Step 3 of the SST procedure described in Section 2, previous SST work has employed the assumption that the annual number of storm counts follows a Poisson distribution, which in turn serves as the basis for the temporal resampling of storms (i.e. for generating the number of storms per year k that will be spatially transposed). RainyDay supports Poisson-based resampling, but also allows the use of an empirical distribution. This distribution is derived from

the number of storms that enter into the storm catalog from each calendar year in the rainfall input dataset. Then, during the temporal resampling step, k is obtained by randomly selecting one of these values. This feature may be useful in regions where storm occurrences exhibit strong clustering (i.e. where there is strong evidence for more storms in some years and fewer in other years for persistent climatological reasons; e.g., Villarini et al., 2013). Section 4.3 examines the impact of this choice on SST results. Other discrete probability distributions, such as the two-parameter negative binomial (Pascal) distribution, can also be used to model clustered count data. RainyDay does not currently use such distributions, since short (typically 10-20 year) remote sensing records may yield poor parameter estimates stemming from the limited number of statistical degrees of freedom.

3.5 "Spin-up" of Initial Conditions

A key issue in the modeling of rainfall driven hazards is to adequately represent initial conditions. In many flood and landslide modeling efforts, the most critical of these initial conditions is antecedent soil moisture, while other states such as river baseflow and water table position may also be relevant. Many hydrologic models allow for the specification of such initial conditions, and thus many design storm-based hazard modeling efforts rely on an assumed soil moisture state, such as a typical or fully saturated condition. Such assumed approaches have previously been used with SST (Wright et al., 2014b), and could be combined with the rainfall scenarios generated via RainyDay. This approach has the downside, however, that the true variability antecedent soil moisture is not captured in hazard predictions. This is particularly important in regions in which heavy rainfall does not necessarily occur in the same season as high soil moisture conditions. A second approach that can capture this variability would be to derive a distribution of antecedent soil moisture from previous long-term (ideally multi-decadal) model simulations. Since there can be substantial variation in how soil moisture is represented in different hazard models, ideally the same model would be used for these long-term simulations and for the hazard scenario modeling. RainyDay offers an alternative option, however, in which initial soil moisture can be "spun up" within the hazard model to represent seasonally realistic initial conditions without the need for long-term simulation.

The spin-up procedure is described for a single rainfall scenario. The month of occurrence of the rainfall scenario is identified based on the "parent storm" that created it. Then RainyDay

identifies the set of X-day periods (where X is a user-defined spin-up period) preceding all parent storms that occur within a user-defined number of months from the date of occurrence of the parent storm. One of those X-day periods is randomly selected and pre-pended to rainfall scenario. This scheme helps to ensure that spin-up conditions are reasonable for the given season. It also helps ensure that spin-up conditions have realistic temporal correlations when pre-pended to the rainfall scenario (for example, if there is a historical tendency for several days of moderate rain prior to heavy storms but several days of heavy rain prior to the main storm doesn't have historical precedent, these conditions will be properly represented). It is important to note, however, that the 10 to 20-year records typical of rainfall remote sensing records may not capture the full variability of "true" initial conditions.

This pre-pending procedure creates rainfall scenario output files that are of $X+t$ day duration. The modeler can then assign an average initial soil moisture condition to initialize each model run, and use the rainfall scenario as input. Soil moisture within the model will then evolve over the spin-up period based on the rainfall (or lack thereof), evapotranspiration, and other model-estimated fluxes. It is important to point out that this spin-up procedure has several limitations. First, it has a storage and computational cost since it can substantially increase the size of the rainfall scenario output files generated by RainyDay and increase the length of each hazard simulation. The importance of these limitations depends on the size of A, the resolution of the input rainfall dataset, and the computational burden of the hazard model. In Section 4.2, for example, we limit X to 6 days, for a total rainfall duration of 10 days. This spin-up period is likely sufficient to spin up moisture in the upper soil layers, but not to fully establish baseflow or deeper groundwater flow. The modeler should evaluate the tradeoffs between longer X and the associated storage and computational costs.

3.6 Parametric Rainfall Intensity

Instead of relying on the rainfall intensity derived from a remote sensing input dataset, a user might prefer to use a parametric distribution to impose rainfall depths on the rainfall output scenarios. RainyDay supports this option. The user can supply a t-hour rainfall depth distribution. This distribution is then applied to the output rainfall scenarios via a normalization procedure that assumes that the supplied distribution corresponds to the annual maximum t-hour rainfall intensity for a single rainfall grid cell. Rainfall space-time structure is still derived from

the remote sensing data. It should be noted, however, that when the resolution of the input remote sensing dataset is coarse relative to the spatial coverage of the rainfall measurement device upon which the parametric distribution is based (for example, the 16-625 km^2 footprint of many satellite rainfall datasets relative to the 0.1 m^2 sampling area of a single rain gage), this approach may be problematic. This procedure is also problematic in regions where such parametric rainfall distributions might be the synthesis of "mixture distributions" of distinct storm types in which rainfall intensity is intrinsically linked to rainfall space-time structure (e.g. Smith et al., 2011), since RainyDay does not distinguish between different storm types. Currently only the three-parameter generalized extreme value distribution (Walshaw, 2013) is supported, though it would be straightforward to add additional choices.

4. Rainfall and Flood Case Studies

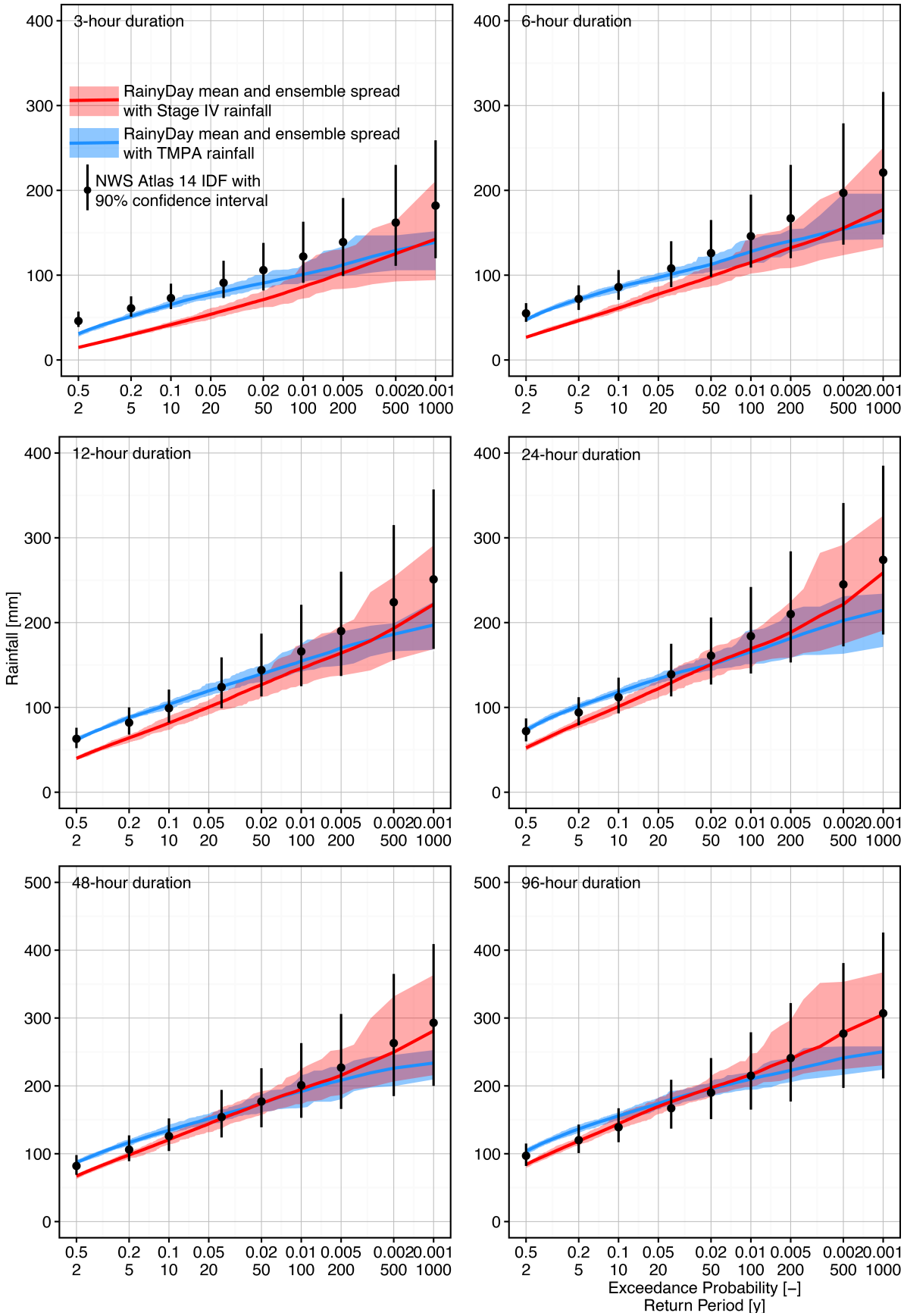
4.1 Rainfall IDF

We generated IDF results for six durations from 3 to 96 hours over a range of p_e between 0.5 and 10^{-3} using RainyDay for single rainfall grid cells in the vicinity of Iowa City, Iowa (Figure 4) using rainfall data from Stage IV (Lin and Mitchell, 2005) and version 7.0 of the Tropical Rainfall Measurement Mission Multi-Satellite Precipitation Analysis (TMPA; Huffman et al., 2010). Stage IV is available through the National Weather Service (NWS) National Center for Environmental Prediction and provides hourly, 4 km resolution rainfall estimates by merging data from the NWS Next-Generation Radar network (NEXRAD; Crum and Alberty, 1993) with rain gages and, in some instances, satellite rainfall estimates. Stage IV has been extensively used in studies of extreme rainfall and flooding. All Stage IV analyses in this paper use data from 2002 to 2014. TMPA merges passive microwave, active radar, and infrared observations from multiple satellites to create a near-global (±50° latitude) rainfall dataset with 3-hourly, 0.25° (approximately 25 km) resolution. Unless otherwise noted, TMPA analyses this study uses the final "research version" of TMPA from 1998-2014, which includes a monthly rain gage-based bias correction. For the results in Figure 4, and most subsequent analyses in this study, *A'* is the rectangular area shown in Figure 3. *A* is set to a single rainfall pixel and each run consists of 100 ensemble members (N=100), producing 100 estimates for each p_e. We compare these results with rain gage-based IDFs from NOAA Atlas 14. Atlas 14 uses L-moment regionalization techniques

534 to combine observations from large number of rain gages. The Atlas 14 analysis for Iowa uses
535 369 rain gages, many of which have records beginning in the late 19th century.
536
537 The range of IDF durations shown in Figure 4 emphasize that RainyDay is flexible in terms of
538 the selection of duration *t*. RainyDay-based IDF estimates using Stage IV exhibit slight
539 systematic underestimation relative to Atlas 14 across a range of p_e except for at the 96-hour
540 duration, where there is a close match. RainyDay-based IDF estimates using TMPA, meanwhile,
541 closely match Atlas 14 for high p_e (except at the 3-hour scale) and underestimates for low p_e for
542 all durations. Underestimation using RainyDay may be attributed to the mismatch in spatial
543 resolution of the remote sensing data (approximately 16 km^2 for Stage IV and 625 km^2 for
544 TMPA) and the rain gages (approximately 0.1 m^2). We have refrained from using ARFs to
545 convert the Atlas 14 point IDF estimates into area-averaged IDFs, since the ARF concept has
546 practical and conceptual limitations (see Wright et al., 2014a). Both the slight overestimation of
547 rainfall depth from TMPA (relative to Stage IV) for more frequent events, and the
548 underestimation for more rare events using both datasets, could potentially be explained by
549 conditional bias (i.e. bias that is dependent on rain-rate; Ciach et al. 2000, see Habib et al., 2009
550 for evidence of conditional biases in TMPA). The convergence between Stage IV-based
551 RainyDay IDFs and Atlas 14 with increasing duration is consistent with both conditional bias
552 and spatial mismatch effects, both of which are known to diminish with increased temporal
553 aggregation. While not definitive, the results in Figure 4 do not clearly point to shortcomings
554 associated with the SST procedure itself.
555
556 In order to highlight both the potential for IDF estimation and probabilistic hazard assessment in
557 data-sparse regions using RainyDay with satellite remote sensing, and some of the associated
558 challenges, we compare 24-hour IDF curves generated using RainyDay for various satellite
559 rainfall datasets for the vicinity of Iowa City (Figure 5). This comparison includes two versions
560 of TMPA: the aforementioned final version which includes monthly rain gage-based bias
561 correction, and TMPA-RT, which is produced in near real-time, does not feature bias correction,
562 and runs from 2000-2014. It also includes two versions of the 30-minute resolution, 8 km
563 Climate Prediction Center (CPC) Morphing Technique (CMORPH; Joyce et al., 2004):
564 CMORPH Corrected, which uses a daily rain gage-based bias correction scheme, and CMORPH

565 Raw, which does not. Finally, it includes the 60-minute, approximately 4 km version of
566 Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks
567 Global Cloud Classification System (PERSIANN-GCCS; Sorooshian et al., 2000), which does
568 not use gage-based bias correction. The results in the top panel of Figure 5 show relatively good
569 agreement between point-scale NOAA Atlas 14 IDFs and single-pixel RainyDay-based IDFs for
570 bias-corrected TMPA and PERSIANN-GCCS, particularly considering the spatial sampling
571 mismatch between the remote sensing data and Atlas 14 mentioned previously, while results
572 based on CMORPH Corrected show systematic underestimation.
573
574 The middle panel of Figure 5 shows how RainyDay can be used to examine the effect of rain
575 gage-based bias correction on satellite-based IDF estimates. In the case of CMORPH, the Raw
576 version overestimates rainfall intensity at all p_e, while results for the Corrected version shows
577 that the daily-scale bias correction scheme seems to overcompensate, leading to systematic
578 underestimation. The TMPA-RT also overestimates at all p_e, though not as severely as
579 CMORPH Raw, while the monthly bias correction scheme used in the final version of TMPA
580 appears to offer superior performance to the daily-scale routine used by CMORPH Corrected. It
581 is not immediately clear why this is the case, particularly since details of the bias correction
582 procedure for CMORPH are not readily available, but relevant considerations include the effect
583 of rainfall detection errors on bias correction (Tian et al., 2007) and the challenge of correcting
584 for conditional biases at short time scales (Wright et al., 2014c). The apparent strong
585 performance of the monthly bias correction is encouraging in the context of Integrated Multi-
586 satellitE Retrievals for GPM (IMERG), a state-of-the-art rainfall dataset that combines various
587 elements from TMPA, CMORPH, and PERSIANN, including TMPA's monthly bias correction
588 (Huffman et al., 2014). The IMERG dataset is not analyzed in this study since the full
589 retrospective dataset is not yet available.
590
591 The bottom panel of Figure 5 shows results similar to those in the top panel, but with A set to a
592 0.5° by 0.5° (approximately 2500 km^2) box centered on Iowa City. The results demonstrate that
593 RainyDay can easily generate spatially aggregated rainfall IDF curves. This is not achievable
594 using standard gage-based IDF curves without the use of ARFs, which, as previously mentioned,

595 have been shown to have limitations. We omit an area-averaged gage-based IDF curve from the
596 bottom panel of Figure 5 for this reason.
597
598 The results shown in Figure 5 (and also Figure 4) have implications for using RainyDay for IDF
599 and hazard estimation in data-sparse regions using satellite remote sensing. First, there can be
600 substantial differences in extreme rainfall estimates between satellite rainfall datasets, and these
601 differences will propagate through to IDF estimates (and to probabilistic hazard estimates, as will
602 be shown in Section 4.2). Furthermore, while comparison with gage-based IDFs (when
603 available) can be used to understand these differences, spatial sampling mismatches complicate
604 comparisons. Findings may not be transferable across regions since the performance of satellite
605 rainfall retrievals vary with region and latitude (e.g. Ebert et al., 2007) and because the quality of
606 the gage-based bias correction schemes that some of satellite datasets employ will vary
607 regionally with the density of rain gage observations that are available.


3-hour duration
6-hour duration
12-hour duration
24-hour duration
48-hour duration
96-hour duration
RainyDay mean and ensemble spread with Stage IV rainfall
RainyDay mean and ensemble spread with TMPA rainfall
NWS Atlas 14 IDF with 90% confidence interval
Rainfall [mm]
Exceedance Probability [–]
Return Period [y]
0.5 0.2 0.1 0.05 0.02 0.01 0.005 0.002 0.001
2 5 10 20 50 100 200 500 1000

608

Figure 4: Comparison of IDF curves from Atlas 14 and RainyDay using the Stage IV and TMPA rainfall datasets for 3-, 6-, 12-, 24-, 48-, and 96-hour durations. Shaded areas for RainyDay estimates denote the ensemble spread. Bars on the NOAA Atlas 14 IDF estimates denote the 90% confidence intervals. Key RainyDay parameters: m=150 storms, A'=[40º to 44º N, 90 º to 96º W].=, A is a single rainfall pixel (approximately 16 km^2 for Stage IV, 625 km^2 for TMPA), N=100, T_{max}=1000. Spatially-uniform transposition and Poisson-based temporal resampling are selected. Stage IV period of record is 2002-2014, TMPA period of record is 1998-2014. Analyses are restricted to April-November period.

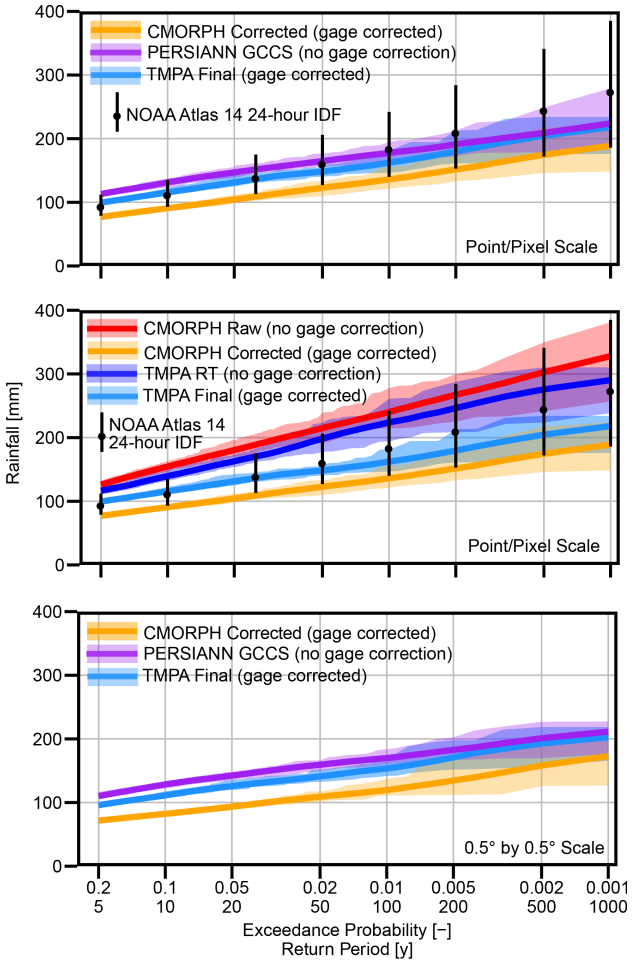


Figure 5: Comparison of IDF curves. Top: 24-hour duration IDF curves at the point scale from NOAA Atlas 14 and at the pixel scale from RainyDay using TMPA Final, PERSIANN-GCCS, and CMORPH Corrected rainfall datasets. Middle: 24-hour duration IDF curves at the point scale from NOAA Atlas 14 and at the pixel scale from RainyDay using TMPA-RT, TMPA Final, CMORPH Raw, and CMORPH Corrected rainfall datasets. Bottom: 24-hour duration IDF curves at the 0.5° by 0.5° scale from RainyDay using TMPA, PERSIANN-GCCS, and CMORPH Corrected rainfall datasets. Shaded areas for RainyDay estimates denote ensemble spread. Bars on the NOAA Atlas 14 IDF estimates denote the 90% confidence intervals. Key RainyDay parameters: m=150 storms, A'=[40º to 44º N, 90 º to 96º W]. A is a single Stage IV rainfall pixel (approximately 625 km^2 for TMPA, 64 km^2 for CMORPH, 16 km^2 for PERSIANN), N=100, T_{max}=1000, t=24

626 **hours . Spatially-uniform transposition and Poisson-based temporal resampling are selected. TMPA Final**
627 **and CMORPH period of record is 1998-2014, TMPA RT period is 2000-2014, PERSIANN GCCS period of**
628 **record is 2004-2014. Analyses are restricted to April-November period.**
629

630 *4.2 Flood Frequency Analysis*

631 In this section, we present flood peak frequency analyses for the 4400 km^2 Turkey River
632 watershed in northeastern Iowa using rainfall scenarios from RainyDay as inputs to the Iowa
633 Flood Center (IFC) Model, a calibration-free distributed hydrologic modeling framework
634 designed primarily for multi-scale flood research and application (see Cunha et al., 2012; Demir
635 and Krajewski, 2013; Mantilla and Gupta, 2005; Moser et al., 2015; Small et al., 2013). Moser et
636 al. (2015) provides a detailed model description and Cunha et al. (2012) performed model
637 validations for flood events in Iowa, showing that the performance of the IFC Model is generally
638 comparable to that of the more heavily-calibrated operational NWS SAC-SMA flood forecast
639 model (Burnash, 1995). The model configuration used here is the same that was used by Moser
640 et al (2015). This study aims only to demonstrate basic features of RainyDay for flood hazard
641 analysis and so does not provide detailed discussion of the IFC Model or comparisons with other
642 available platforms. For a discussion of the value of calibration-free, distributed hydrologic
643 models for multi-scale flood modeling, the reader is directed to Wright et al. (2014b) and, in
644 particular, Cunha et al. (2012). The full multi-scale hazard estimation capabilities of SST and
645 RainyDay can, in principle, be harnessed using any distributed hydrologic or mass wasting
646 model, while some of the capabilities can be achieved through the use of lumped models.
647

648 A limited set of model hydrograph validation is provided in Figure 6 for the 2008 and 2014
649 April-July periods, during which major flooding occurred throughout Iowa (see Smith et al.,
650 2013 for a detailed examination of the hydrometeorology of the 2008 flood season). The model
651 is run both with Stage IV and the final (gage-corrected) version of TMPA rainfall. Comparisons
652 with U.S. Geological Survey (USGS) stream gage observations are provided at four locations,
653 with upstream drainage areas ranging from 900 to 4000 km^2. All hydrographs are normalized by
654 the median annual flood (p_e=0.5) to facilitate comparison across watershed scales. Median
655 annual flood estimates are taken from the USGS StreamStats system
656 (http://water.usgs.gov/osw/streamstats/). Model performance varies from event to event but there

is no clear evidence of systematic bias in the streamflow predictions as a function of event magnitude or drainage area. Predictions based on Stage IV are generally better than TMPA, and in fact several time periods show serious problems with the timing of TMPA-based simulations. In the 2008 flood season, TMPA incorrectly identifies the late April event as the largest for that year, rather than the early June floods.

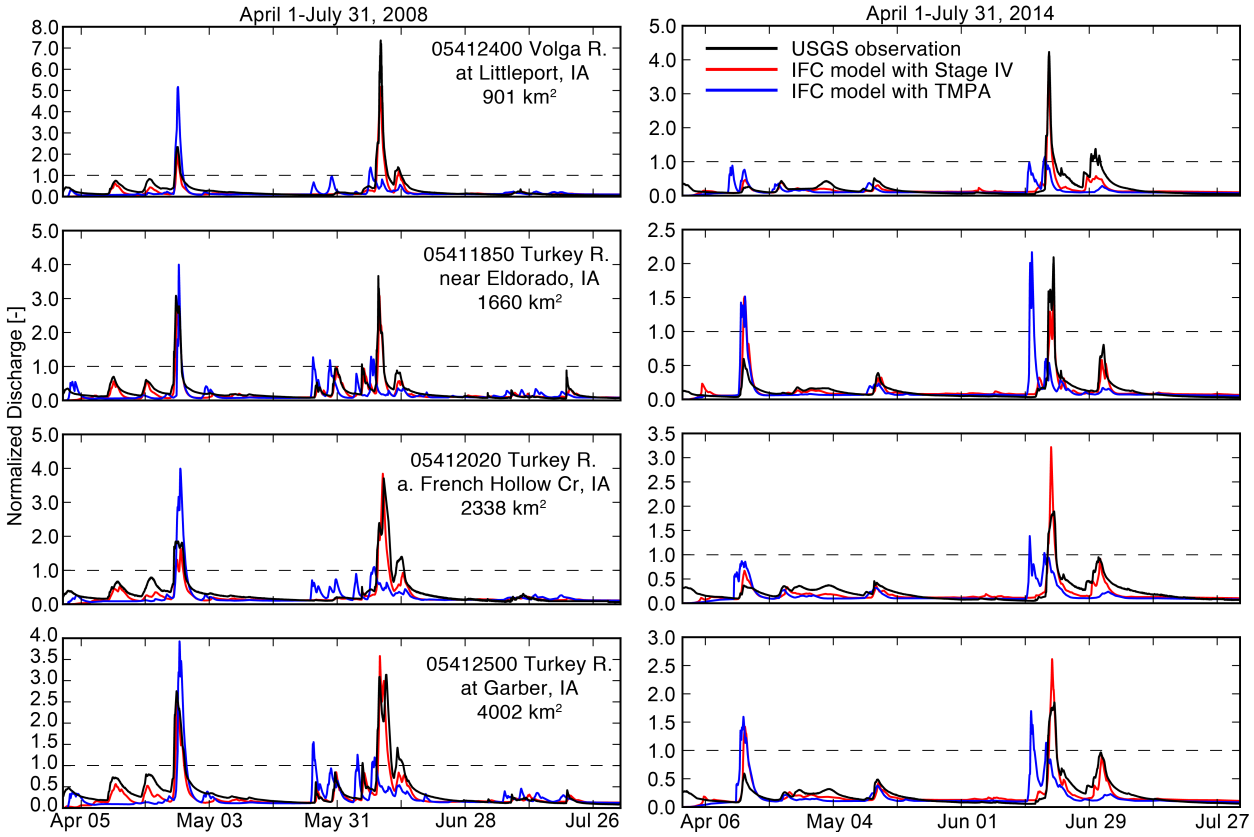


Figure 6: IFC model validation for 2008 and 2014 flood seasons (left and right panels, respectively) at four USGS stream gaging sites. Hydrographs are normalized by the median annual flood, which is indicated by dashed horizontal lines.

We compare observed and simulated flood peaks for the 2008-2014 April-November period (Figure 7). All observed flood peaks that exceed 100 $m^3 s^{-1}$ are extracted from the four USGS stream gaging records. Then the corresponding flood peaks predicted by the IFC model are extracted from simulated hydrographs based on Stage IV and TMPA rainfall (left panel and right panels of Figure 7, respectively). To allow for modest errors in flood peak timing, a window of 48 hours centered around the observed peak is used to identify the corresponding simulated peaks. All peaks in Figure 7 are normalized by the median annual flood for to facilitate

comparison across basin scales. As a rule of thumb, peaks below the median annual flood can be considered "within bank," while peaks above the median annual flood can be considered "out-of-bank," meaning the flood magnitude is large enough to exceed the normal confines of the river channel and spill into the floodplain. The left panel of Figure 7 shows that, while there is modest scatter in the Stage IV-based flood peak simulations, there is no obvious systematic bias with watershed scale or event magnitude. The TMPA-based simulations in the right panel of Figure 7 exhibit greater scatter, generally poor performance, and show some low bias across a range of event magnitudes. While not an exhaustive, the validation shown in Figures 6 and 7 suggests that streamflow prediction accuracy in the IFC model is driven primarily by the accuracy of the input rainfall rather than by model structure, consistent with Cunha et al. (2012), and that the limited accuracy of satellite rainfall inputs, even with gage-based bias correction, can translate into relatively poor model performance.

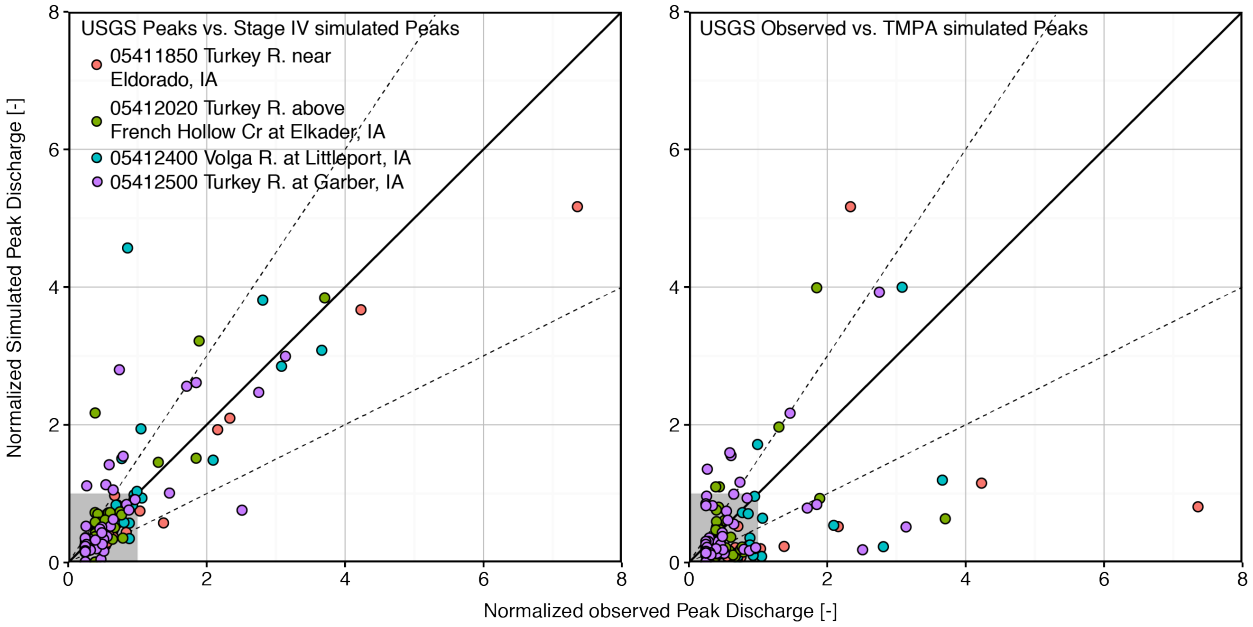


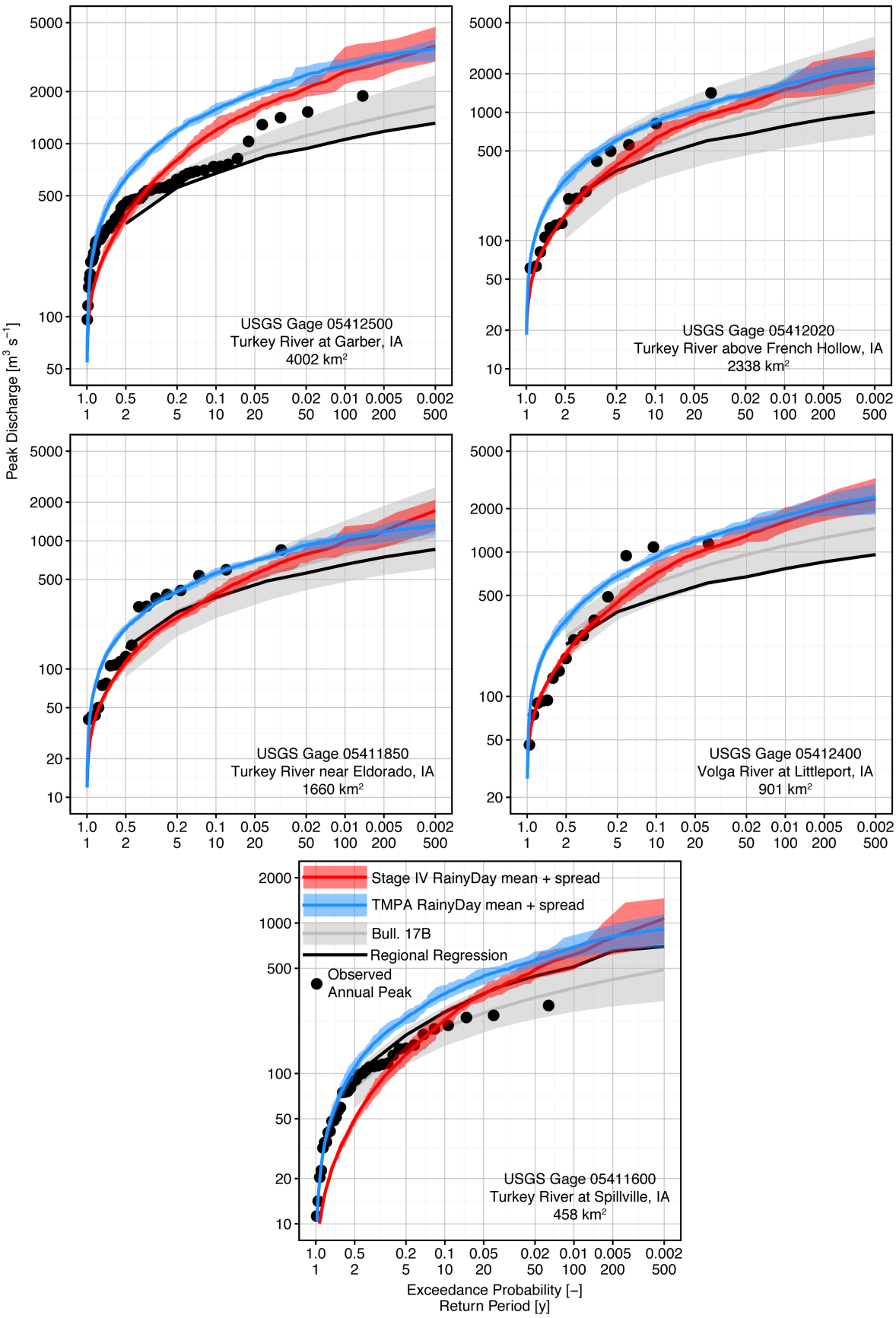
Figure 7: Peak discharge validation for 2008-2014 April-November period at four USGS stream gaging stations. All events for which the USGS observations exceeded 100 $m^3 s^{-1}$ are shown, and peak discharges are normalized by the median annual flood. Simulated peaks using the IFC model with Stage IV (TMPA Final) rainfall inputs are compared with USGS observed peaks in the left (right) panel. Straight black lines indicate 1:1 correspondence, while dashed lines indicate the envelope within which the modeled values are within 50% of observed. Grey boxes in the lower lefthand corners of each panel highlight all events less than the median annual flood.

696 We performed IFC model simulations using RainyDay rainfall scenarios developed from both
697 the Stage IV and final gage-corrected TMPA rainfall datasets. For each rainfall dataset, we ran
698 ten ensemble members (i.e. N=10), each consisting of 500 rainfall scenarios (i.e. T_{max}=500). At
699 any point along the modeled river system, therefore, flood peak exceedance probabilities as low
700 as 0.002 (500-year return period) could be directly derived from the IFC Model predictions. The
701 Stage IV and TMPA-based storm catalogs for the Turkey River include 150 storms, drawn from
702 the April-November rainfall record (2002-2014 for Stage IV, 1998-2014 for TMPA). *A'* is an
703 area covering most of Iowa, southwestern Wisconsin, and southeastern Minnesota in the United
704 States. t = 96 hours for all simulations in this section.
705
706 We initialize each simulation with a spatially uniform initial soil moisture value found to be
707 typical for the region. Rainfall from a seasonally-based six-day "spin-up" period was then
708 prepended to each 96-hour storm period as per Section 3.5, for a total rainfall input time period
709 of ten days. Spatial variations in both soil moisture and river flow were therefore allowed to
710 develop in each simulation prior to the arrival of the main storm. It should be noted that
711 restricting the rainfall record to April-November, in addition to the lack of snowfall functionality
712 in RainyDay and snowpack functionality in the IFC model, means that snowmelt-driven flooding
713 is not considered in the analyses. In Iowa, snowmelt is generally a minor though non-negligible
714 flood mechanism. We do not evaluate the accuracy of these spin-up soil moisture and river flow,
715 and in fact such evaluation is relatively challenging due to the paucity of long-term soil moisture
716 observation records that would be needed to correlate with river flow. As discussed in Section
717 3.2, SST and RainyDay facilitates "decoupling" of discharge p_e from rainfall p_e. Though not
718 demonstrated explicitly, this decoupling is reflected in the RainyDay-based frequency analyses
719 in this section, in that the role of spun-up initial conditions and rainfall space-time structures can
720 produce discharge p_e that are different from the p_e of the input rainfall scenarios.
721
722 RainyDay-based frequency analysis results are shown for five subwatersheds of the Turkey
723 River, ranging in drainage area from approximately 460 to 4000 km^2 (Figure 8). Also included in
724 Figure 8 are two types of frequency analyses derived from USGS stream gage observations and
725 taken from Eash et al. (2013) and retrieved from the USGS StreamStats system. The first is
726 developed using standardized methods described in Bulletin 17B (Interagency Advisory

727 Committee on Water Data, 1982) using the log-Pearson Type III distribution (henceforth referred
728 to as the LP3 distribution) with a regionalized skew coefficient. The second is based on regional
729 regression equations that consider drainage basin area and shape as well as some soil and
730 geological properties. Eash et al. (2013) report 121 years of data for Turkey River at Garber, near
731 Eldorado, and above French Hollow, while 63 years are reported for Volga River at Littleport
732 and 45 years for Turkey River at Spillville. It should be noted that these record lengths refer to
733 "historic record length" described in Section V.B.10 of Bulletin 17B and do not correspond to
734 length of the USGS annual maxima streamflow timeseries available on the USGS National
735 Water Information System (http://nwis.waterdata.usgs.gov/nwis), which are much shorter. All
736 available USGS streamflow observations for the five sites are also shown, where p_e is estimated
737 using the Cunnane plotting position (Cunnane, 1978; $p_e^i = [i - 0.4] / [X + 0.2]$, where i is the rank
738 of the observation and X is the number of observations). Other common plotting position
739 formulae produce similar results (not shown) and do not alter the conclusions that follow.
740
741 For all five locations shown in Figure 8, the SST-based peak discharge estimates using TMPA
742 are higher than those using Stage IV for $p_e<0.01$, generally converging toward the Stage IV
743 estimates as p_e decreases, and in some cases yielding lower estimates for p_e less than about 0.005.
744 This is consistent with the rainfall IDF results from RainyDay shown in Figure 4 and are
745 suggestive of conditional biases in the TMPA dataset. This in indeed confirmed in Figure 9,
746 which shows watershed-specific IDF curves for the entire Turkey River watershed from
747 RainyDay using TMPA and Stage IV. The USGS streamflow observations shown in Figure 8
748 agree reasonably well with the Stage IV-based estimates for $p_e>0.5$, with the exception of the
749 smallest subwatershed, Turkey River at Spillville, where Stage IV produces low peak estimates.
750 For $p_e<0.5$, there is a lack of consistency. For example, Turkey River at Garber shows higher
751 estimates from Stage IV than the streamflow observations, while the reverse is true for Turkey
752 River at French Hollow and near Eldorado. Deviations from the USGS observations do not show
753 a systematic scale dependency.
754
755 Both RainyDay-based frequency analyses and the USGS streamflow observations are generally
756 higher than the USGS frequency analyses for p_e less than about 0.2. One exception is the set of
757 USGS observations for Turkey River at Spillville, which are lower than both the RainyDay

758 estimates and the regional regression but generally consistent with the Bulletin 17B analysis. The
759 regional regression results for Turkey River at Spillville are greater than the USGS regionalized
760 LP3 estimates, while the reverse is true for the four larger subwatersheds. Interestingly, some of
761 the USGS observations fall outside of the 90% confidence intervals of the LP3 analyses for
762 Turkey River near Eldorado, Volga River at Littleport, and Turkey River at Garber. In the case
763 of the latter station, the five most intense floods are near or above the upper 95% confidence
764 bound, a finding that is explored in more detail in the following paragraphs.
765
766 It should be noted that with the exception of Turkey River at Garber, the differences between the
767 RainyDay-based analyses are roughly similar in magnitude to the differences between the two
768 different USGS approaches. This fact, along with the underestimation shown by USGS
769 frequency analyses relative to the USGS peak discharge observations at several sites, suggests
770 that the RainyDay-based frequency analyses should not be dismissed out of hand as being too
771 high for low p_e. In fact, as the next example shows, there is observational evidence that supports
772 the validity of the RainyDay-based results in light of possible nonstationarity in flooding. It
773 should be noted that discharge-based frequency analyses, even in stationary situations with long
774 records, are not necessarily superior to hydrologic modeling methods. Analyses by Smith et al.
775 (2013) suggest that peak discharge measurement errors may be substantial for a recent major
776 flood events in Iowa. The propagation of discharge measurement errors through frequency
777 analysis is poorly understood (e.g., Petersen-Øverleir and Reitan, 2009; Petersen-Øverleir, 2004;
778 Potter and Walker, 1985). Rogger et al. (2012) reported significant differences between two
779 commonly-used flood frequency analysis approaches for ten small alpine watersheds in Austria,
780 one based on a stream gage-based statistical method and the other on design storm methods
781 combined with a hydrologic model. The latter method produced higher discharge values than the
782 former, and the authors discuss possible explanations and deficiencies in both approaches while
783 concluding that in at least some situations, hydrologic modeling using rainfall inputs will
784 produce superior results.
785
786 Of the five USGS stream gage locations shown in Figure 8, only the gage at Garber, Iowa has a
787 long (82-year), unbroken annual peak discharge record. We use this record to better understand
788 the discrepancies between the RainyDay-based results and the USGS frequency analyses from

789 Eash et al. (2013), and in particular to contrast the methods in the context of potential
790 nonstationarity in flood processes. The top panel of Figure 10 shows that the same RainyDay and
791 USGS frequency analyses shown in Figure 8 for Turkey River at Garber. In this case, however,
792 the USGS observations have been divided into two groups; one for all peaks occurring from
793 1933 to 1989, and the second for all peaks occurring from 1990 to 2014. The plotting position-
794 based p_e is recalculated for each group of observations. The 1933-1989 subgroup shows higher
795 discharges than either RainyDay Stage IV or USGS discharges for p_e>0.5, and lower discharges
796 for p_e less then about 0.2. The 1990-2014 subgroup, meanwhile, matches closely with the
797 RainyDay-based frequency analyses with Stage IV.
798


Peak Discharge [m^3 s^{-1}]
USGS Gage 05412500
Turkey River at Garber, IA
4002 km^2
USGS Gage 05412020
Turkey River above French Hollow, IA
2338 km^2
USGS Gage 05411850
Turkey River near Eldorado, IA
1660 km^2
USGS Gage 05412400
Volga River at Littleport, IA
901 km^2
USGS Gage 05411600
Turkey River at Spillville, IA
458 km^2
Stage IV RainyDay mean + spread
TMPA RainyDay mean + spread
Bull. 17B
Regional Regression
Observed
Annual Peak
Exceedance Probability [–]
Return Period [y]

799

Figure 8: Peak discharge analyses using RainyDay with Stage IV and TMPA rainfall remote sensing data and the IFC Model, compared against USGS stream gage-based analyses for five subwatersheds of the Turkey River in northeastern Iowa. Shaded areas for RainyDay estimates denote the ensemble spread. Bars on the USGS Bulletin 17B estimates denote the 90% confidence intervals. Confidence intervals are not available for the USGS regional regression. Key RainyDay parameters: m=150 storms, A'=[40º to 44º N, 90 º to 96º W], A is the watershed upstream of the USGS streamgage at Garber, IA, N=10, T_{max}=500, t=96 hours. Spatially-uniform transposition and Poisson-based temporal resampling are selected. Stage IV period of record is 2002-2014, TMPA period of record is 1998-2014. RainyDay Analyses are restricted to April-November period.

Taken together, this suggests a regime shift toward more extreme flooding since 1990 and a reduction in the magnitude of more average floods. Evidence of this regime shift can be seen in the annual peak time series in the bottom panel of Figure 10. We fit a nonparametric linear regression to the 1933-2014 time series using the nonparametric Theil-Sen estimator (Sen, 1968) and a statistically significant (p-value<0.05) downward trend was found. In contrast, using ordinary least squares, an insignificant upward trend is found over the same period. Thus when the influence of the most extreme values is minimized through nonparametric statistical methods, there is a tendency toward smaller flood peaks over time that is not evident with parametric methods, which are more sensitive to the recent extremes.

The top panel of Figure 10 shows that the period of apparent elevated flood activity is well captured by RainyDay, while the preceding period is not, presumably because the IFC model reflects recent land use changes and because the input rainfall data are relatively recent. In general, whether or not this constitutes a strength or limitation of RainyDay depends on the underlying causation of nonstationary flood activity. If flood nonstationarity results from a climate-driven secular trend in extreme rainfall, then the results from RainyDay using relatively short and recent rainfall remote sensing records should be understood as more "up-to-date" estimates of flood frequency compared to approaches, such as the USGS analyses, that use longer stream gage or rain gage records. The same is true if there is a secular trend in flooding due to urbanization or other land-use changes, so long as these changes are properly incorporated into the hydrologic model. In the case of Iowa, flooding has been shown to be affected by land-use change (Villarini and Strong, 2014) and by climate change (Mallakpour and Villarini, 2015). If, on the other hand, flood or rainfall nonstationarity has a periodic structure due to a slowly-

varying climate mode, then the results from SST may only adequately reflect the true flood frequency for the phase of the mode that overlaps with the remote sensing record. It should also be recognized that a period of higher or lower flood activity at a particular location could result from pure randomness (i.e. in absence of both secular and periodic trends). SST should be relatively robust to this possibility through the sampling storms from a larger region.

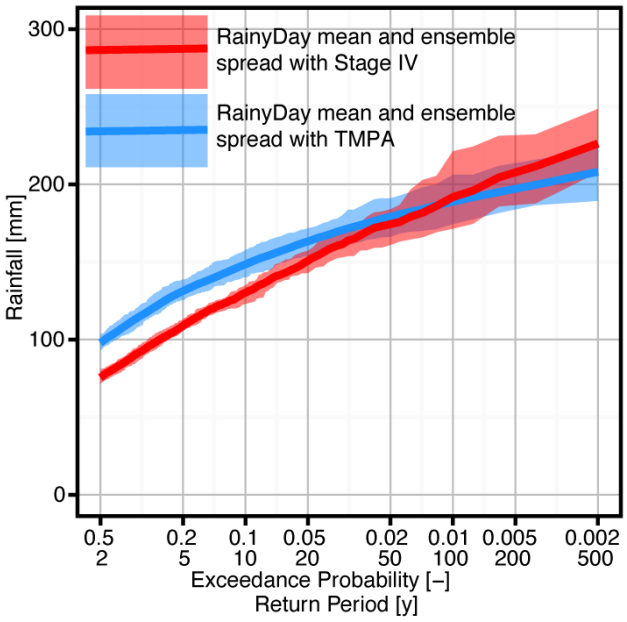
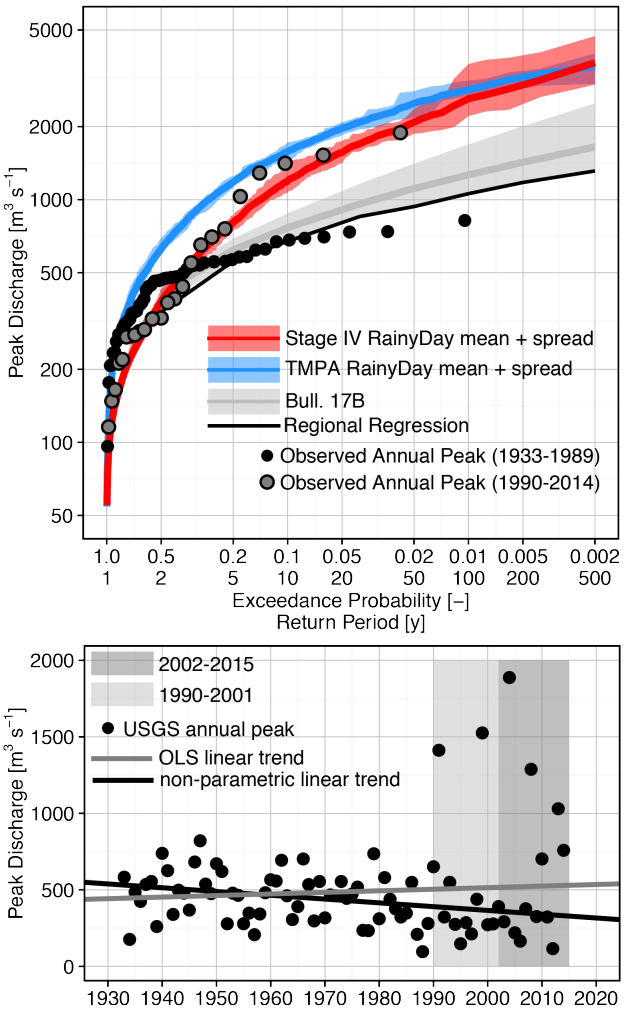


Figure 9: IDF analyses for Turkey River using RainyDay with Stage IV and TMPA rainfall remote sensing data. Shaded areas for RainyDay estimates denote the ensemble spread. Key RainyDay parameters: m=150 storms, A' = [40° to 44° N, 90 ° to 96° W], A is the 4400 km^2 watershed upstream of the confluence with the Mississippi River. N=100, T_{max}=500, t=96 hours, and spatially-uniform transposition and Poisson-based temporal resampling are selected. Stage IV period of record is 2002-2014, TMPA period of record is 1998-2014. Analyses are restricted to April-November period.



847

848 **Figure 10: Top panel—four peak discharge analyses for the location of the USGS stream gage at Garber, IA:**
849 **RainyDay with Stage IV and TMPA rainfall and USGS frequency analyses using regional regression**
850 **relationships and Bulletin 17B methods. Shaded areas for RainyDay estimates denote the ensemble spread.**
851 **Bars for the Bulletin 17B-based analysis denote the 90% confidence intervals. Confidence intervals are not**
852 **available for the USGS regional regression. Bottom panel—annual peak discharge time series for 1932-2014**
853 **for the Garber gage. Linear trend lines in the bottom panel use non-parametric Thiel-Sen regression (Sen,**
854 **1968) and ordinary least squares (OLS). Key RainyDay parameters: *m*=150 storms, *A'* = [40º to 44º N, 90 º to**
855 **96º W], *A* is the watershed upstream of the USGS streamgage at Garber, IA, *N*=10, T_{max}=500, *t*=96 hours.**
856 **Spatially-uniform transposition and Poisson-based temporal resampling are selected. Stage IV period of**
857 **record is 2002-2014, TMPA period of record is 1998-2014. RainyDay Analyses are restricted to April-**
858 **November period.**

859

860

4.3 SST Sensitivity to Record Length and User-defined Parameters

In this section, we examine the sensitivity of SST to the length of the input dataset and to different user-defined parameters and options introduced in Sections 2 and 3. Specific topics that are examined include the optional non-uniform spatial transposition (Section 3.3), empirically-based temporal resampling (Section 3.4) and the size of the transposition domain *A'*. In all cases, it should be kept in mind that the specific results pertain to the Iowa study area and may not be generalizable to other locations. The intention is to demonstrate some important concepts and pitfalls associated with RainyDay, and provide a possible framework for assessing performance in different locations and applications.

The core concept behind SST is "space-for-time substitution," in which storms over a larger region help to inform estimates of rare rainfall in a particular subregion. A common critique of coupling SST with rainfall remote sensing datasets is that such data records are relatively short (approximately 10 to 20 years at time of writing) and thus may not contain sufficient numbers of extreme events at the regional scale to leverage this substitution property and accurately recreate the properties of rare rainfall events. To examine this critique, we turn to a longer dataset: CPC-Unified, a daily rain gage-based gridded rainfall dataset that has a spatial resolution of 0.25° over the conterminous United States (Chen et al., 2008; Xie et al., 2007). Though the spatial and temporal resolution of CPC-Unified is generally insufficient for fine-scale flood modeling, its long record—1948 to present—makes it ideal for evaluating the sensitivity of SST-based IDF estimates to record lengthWe examined several stationarity measures over the transposition domain *A'* (which, as in Section 4.1, roughly encompasses the state of Iowa), including the average number of storm counts per year and the mean, median, and standard deviation of storm rainfall depth (results not shown). None of these measures revealed significant temporal trends, generally consistent with Villarini et al. (2011). This may contradict the apparent flood nonstationarity in the Turkey River watershed discussed in Section 4.2, or may point to land-use change as the predominant source of non-stationarity in Turkey River, but rigorous examination is beyond the scope of this paper.

We use a bootstrapping approach to examine variability in IDF estimates derived from the CPC-Unified data using RainyDay and how this variability evolves as the length of the record

increases. All IDF estimates in this section are for 1-day rainfall over averaged over a 0.5° by 0.5° box. We generate *n*-year long input rainfall datasets by randomly selecting *n* years of CPC-Unified data without replacement from the 1948-2014 period. Each of these datasets is then used as the basis for a single run of RainyDay with 100 ensemble members and with $m = 10n$ (leading to λ=10 storms per year). We repeat this procedure to create 25 datasets for each value of n = 10, 20, 30, 40, 50 years.

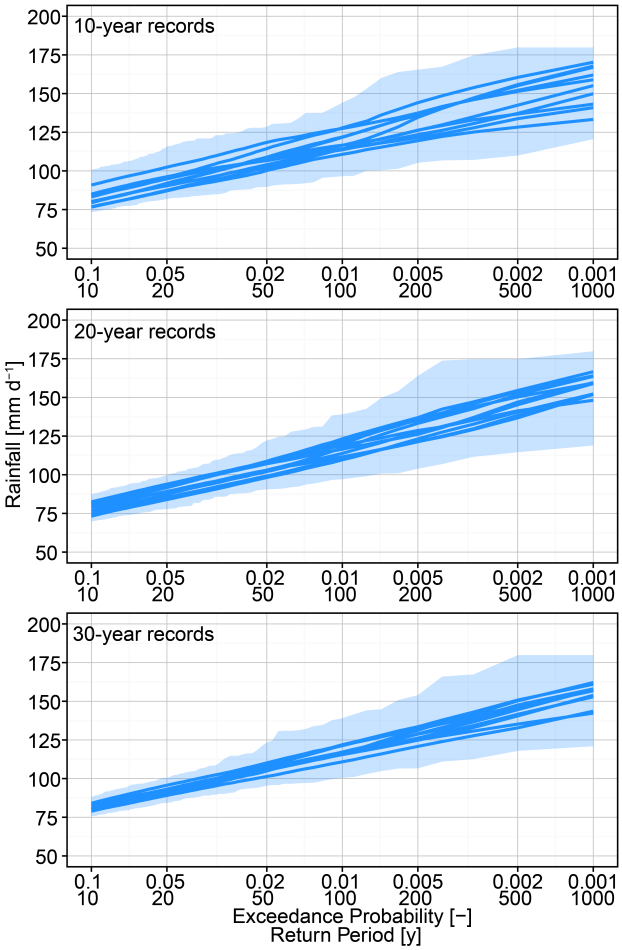


Figure 11: The effect of the rainfall record length on daily rainfall IDF curves estimated using RainyDay with the CPC-Unified daily rainfall over Iowa, United States. Each panel shows the ensemble mean (solid lines) for ten independent runs of RainyDay. The shaded areas denote the maximum spread across the ten runs. Key RainyDay parameters: m=10n storms (where n varies by specified record length), A'=[40° to 44° N, 90 ° to 96° W], A is a 0.5° by 0.5° box, N=100, T_{max}=1000, t=1 day, spatially-uniform transposition and Poisson-based temporal resampling. Analyses are restricted to April-November period.

Substantially more variability is evident in the ensemble mean and spread of the IDF estimates using 10 years of CPC-Unified data than using 20 years, while change in variability is negligible

between runs using 20 years and 30 years of data (Figure 11). We also examined the variability of relative deviations in the ensemble IDF means, minima, and maxima from RainyDay between the *n*-year runs and IDFs based on the full 67-year dataset (Figure 12). The boxplots show that the majority of the deviations in the *n*-year IDF ensemble means, minima, and maxima are less than 10% and that the vast majority are less than 20% for any given p_e. For most p_e, there are substantial reductions in deviation when the records increase in length from n = 10 to n = 20 years. The reductions in deviation are less when the record length increases beyond 20 years.

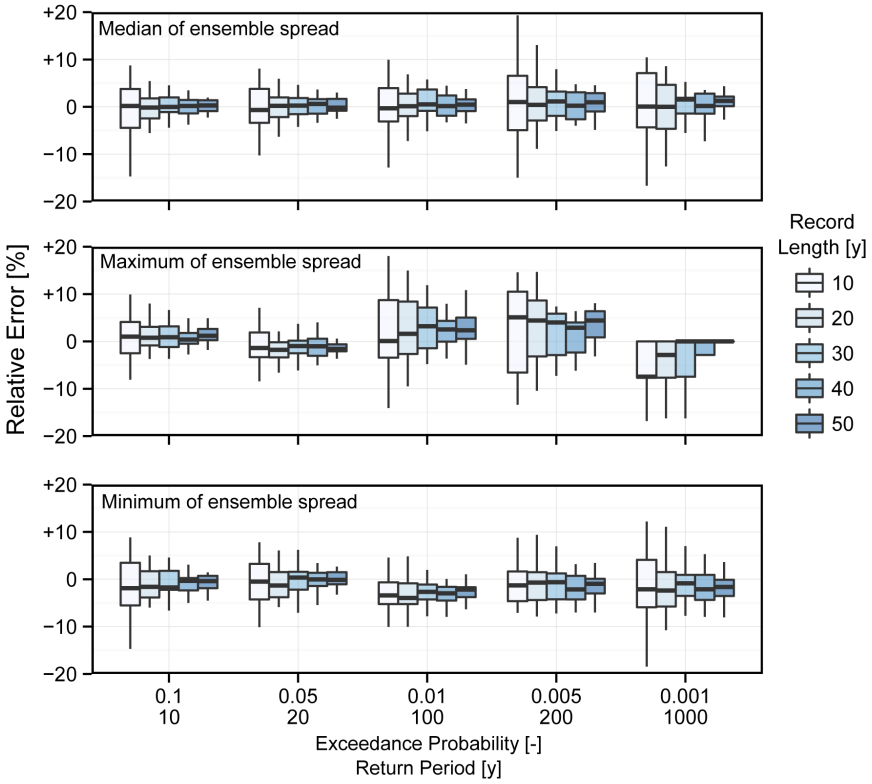


Figure 12: The effect of rainfall record length on variability in daily rainfall IDF estimated using RainyDay with CPC-Unified data over Iowa, United States for 0.1, 0.05, 0.01, 0.005, and 0.001 exceedance probabilities. Each boxplot shows the variability of a particular rainfall quantity at a given exceedance probability across 25 independent runs of RainyDay. Specific rainfall quantities shown are the ensemble mean (top panel), ensemble maximum (middle panel), and ensemble minimum (bottom panel). Boxes denote the lower and upper quartiles and whiskers indicate the extent of the +/-1.5 interquartile range. Key RainyDay parameters: *m*=10*n* storms (where *n* varies by specified record length), *A*'=[40º to 44º N, 90 º to 96º W], *A* is a 0.5° by 0.5° box, *N*=100, T_{max}=1000, *t*=1 day, spatially-uniform transposition and Poisson-based temporal resampling. Analyses are restricted to April-November period.

926 Unless the intensity of the rainfall inputs is perturbed stochastically, SST-based frequency
927 analyses have an inherent upper bound. This upper bound corresponds to the most intense
928 rainstorm in the storm catalog transposed in such a way that rainfall over A is maximized. The
929 lack of positive deviations in the ensemble maxima at $p_e = 10^{-3}$ (middle panel of Figure 12; also
930 in certain realizations shown in Figure 11) show where the SST procedure "encounters" this
931 upper limit.
932
933 While the results in this section are by no means exhaustive and the conclusions are specific to
934 the Iowa study region and could vary in different physiographic regions, they nonetheless
935 suggest that concerns over the use of relatively short remote sensing records with SST may be
936 overstated and that remote sensing datasets, many of which are approaching 20 years in length,
937 should provide relatively robust estimates that will improve as these datasets continue to grow in
938 length. This emphasizes the fact that rainfall events that would be considered rare from the
939 perspective of a single location or watershed can occur relatively frequently from a regional
940 perspective. This is qualitatively consistent with the findings of Troutman and Karlinger (2003),
941 who estimate that a flood with $p_e > 10^{-2}$ occurs on average every 4.5 years at at least one of the
942 193 USGS stream gage sites in their Puget Sound study region.
943
944 A potentially important issue related to short data records in SST, previously mentioned in
945 Section 3.3, can arise if, instead of assuming that the probability of storm occurrence is uniform
946 across the transposition domain, non-uniform spatial transposition is used instead (such as the
947 approach used in Wilson and Foufoula-Georgiou, 1990 or the optional scheme in RainyDay
948 described in Section 3.3). Using the bootstrapping approach with the CPC-Unified dataset
949 described above, visual inspection of storm probability-of-occurrence maps such as the one
950 shown in Figure 3 reveal that there can be substantial variations in the spatial distribution of
951 historical storms when rainfall records are short (results not shown). These variations tend to
952 diminish as the length of record increases, as do their impacts on IDF estimates. More variation
953 is evident in the median IDFs from ten independent runs of RainyDay, for example, using non-
954 uniform transposition than using uniform transposition when n=10 years (Figure 13, left panels).
955 When using non-uniform transposition, variability diminishes when n=20 years and a systematic
956 increase in rainfall intensity for $p_e > 0.02$, relative to the uniform transposition case, emerges

(Figure 13, right panels). Given these results, we recommend that the assumption of uniform transposition be used in the absence of strong physically-based reasoning and observational support for non-uniform transposition. It is possible, however, that this explains the IDF underestimation by RainyDay with Stage IV for high p_e relative to Atlas 14 shown in Figure 4, where uniform spatial transposition was used.

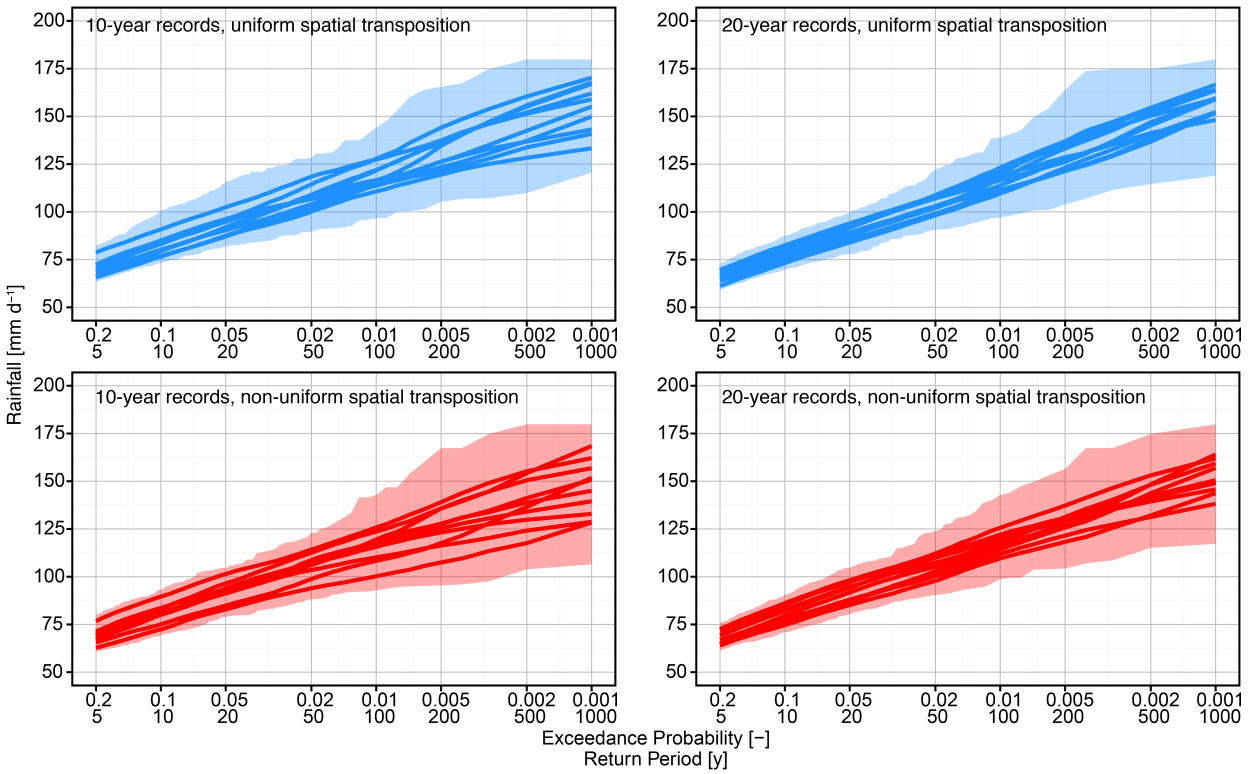


Figure 13: The effect of the spatial transposition scheme on daily rainfall IDF curves estimated using RainyDay with the CPC-Unified daily rainfall over Iowa, United States. Each panel shows the ensemble mean (solid lines) for ten independent runs of RainyDay. The shaded areas denote the maximum spread across the ten runs. The specific years that comprise the input dataset vary. Key RainyDay parameters: m=10n storms (where n varies by specified record length), A'=[40° to 44° N, 90 ° to 96° W], A is a 0.5° by 0.5° box, N=100, T_{max}=1000, t=1 day. Poisson-based temporal resampling is used. Analyses are restricted to April-November period.

As mentioned previously, RainyDay supports either the Poisson-based resampling that has traditionally been used with SST, or an empirical scheme described in Section 3.4. There do not appear to be substantial systematic differences between the results from RainyDay using these two schemes with 10-year records (Figure 14, left panels), but, similar to Figure 13, when 20-year records are used, there is a tendency toward higher rainfall estimates for p_e>0.02. Results

may differ in other regions, particularly where temporal clustering of storms is very strong or where rainstorms are very infrequent. It is recommended that the modeler assess clustering using an independent long-term rainfall data source if available, in addition to assessing sensitivity to this option in RainyDay. As with the spatial transposition schemes, the choice of temporal resampling scheme does not appear to have a substantial impact on low p_e estimates.

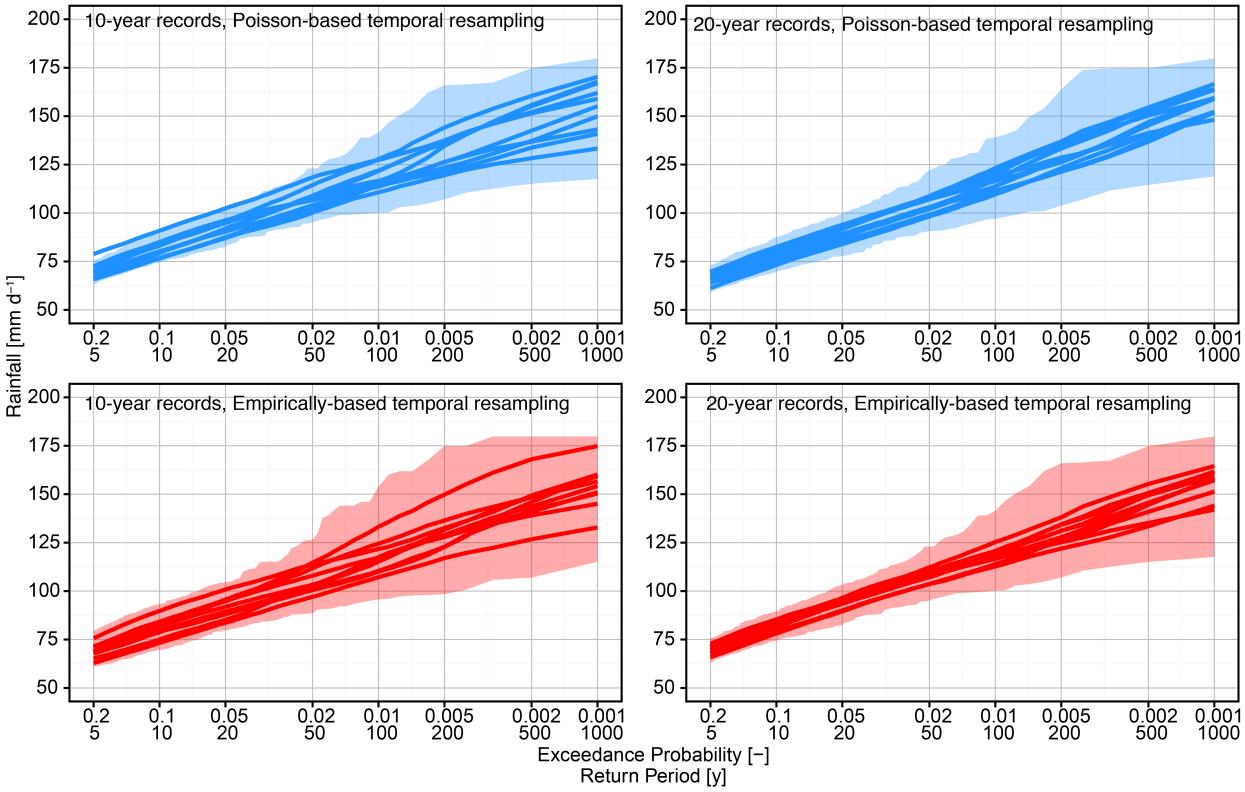
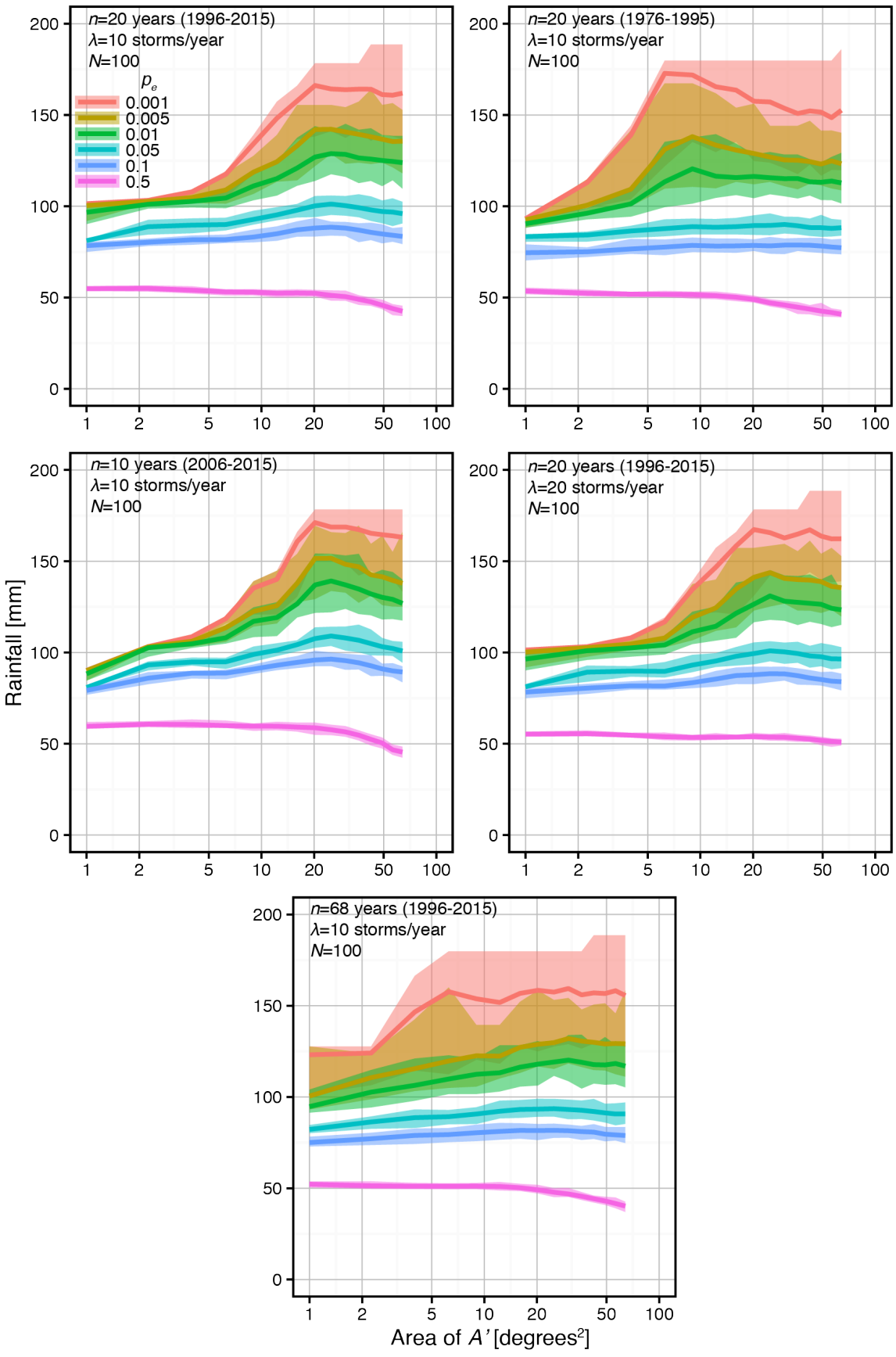


Figure 14: The effect of the temporal resampling scheme on daily rainfall IDF curves estimated using RainyDay with the CPC-Unified daily rainfall over Iowa, United States. Each panel shows the ensemble mean (solid lines) for ten independent runs of RainyDay. The shaded areas denote the maximum spread across the ten runs. The specific years that comprise the input dataset vary. Key RainyDay parameters: m=10n storms (where n varies by specified record length), A'=[40° to 44° N, 90 ° to 96° W], A is a 0.5° by 0.5° box, N=100, T_{max}=1000, t=1 day. Spatially uniform transposition is used. Analyses are restricted to April-November period.

We also examine the sensitivity of RainyDay results to the size of A' (Figure 15). To do so, we run RainyDay for various square domains ranging from 1° by 1° up to 10° by 10°, while holding A fixed at a 0.5° by 0.5° box. Then the evolution of rainfall intensity is examined for a range of p_e as a function of A'. This is repeated for a several different record lengths and for two values of

λ. Interestingly, while there is a general tendency for intensity estimates to stabilize as *A'* grows, the behavior is not asymptotic (though roughly so for n=68 years). The high exceedance probability estimates (p_e=0.5) tend to be stable over a large range of *A'* and then descrease for very large values, due to the tendency for synthetic years to be created in which no storm is transposed directly over *A*. This is the root of potential low biases mentioned in Step 2 of the SST procedure described in Section 2. However, Figure 15 demonstrates that this tendency for a decrease in intensity estimates for large *A'* extends to smaller p_e values as well, and that there is a critical value of *A'* at which the estimated intensity is roughly maximized. This critical value appears to vary more by the particular period of record than by the length of record. For example, the 20-year record from 1976-1995 yielded a critical value of *A'* that is lower than the critical value from 20-year record from 1996-2015. This points to the fact that the existence and number of major storms within *A'* during the record period is very important (Wright et al., 2014b reached the same conclusion through different means).

These results also indicate that increasing m (thus increasing λ) can mitigate the reduction in estimated intensity for values of *A'* larger than the critical value. This result suggests that, if the modeler is interested in hazard estimation across a range of p_e, he or she should choose a relatively large m. A diagnostic framework within the RainyDay software to identify this critical value of *A'* for a given value of m (or vice versa) for different p_e would be useful but does not currently exist.


n=20 years (1996-2015)
λ=10 storms/year
N=100
p_e
0.001
0.005
0.01
0.05
0.1
0.5
n=20 years (1976-1995)
λ=10 storms/year
N=100
n=10 years (2006-2015)
λ=10 storms/year
N=100
n=20 years (1996-2015)
λ=20 storms/year
N=100
n=68 years (1996-2015)
λ=10 storms/year
N=100
Rainfall [mm]
Area of A' [degrees²]
0
50
100
150
200
1
2
5
10
20
50
100

1015

Figure 15: The effect of the size of the transposition domain *A'* on daily rainfall IDF curves estimated using RainyDay with the CPC-Unified daily rainfall over Iowa, United States using a range of record lengths. Key RainyDay parameters: *m*=10*n* storms (where *n* varies by specified record length), *A'* is a square of varying size, *A* is a 0.5° by 0.5° box, *N*=100, T_{max}=1000, *t*=1 day, spatially-uniform transposition and Poisson-based temporal resampling. Analyses are restricted to April-November period.

5. Discussion and Conclusions

In this paper we introduce RainyDay, a Python-based platform that couples rainfall remote sensing data with a technique known as Stochastic Storm Transposition (SST) that effectively "lengthens" the extreme rainfall record through temporal resampling and spatial transposition of observed rainstorms. It produces probabilistic extreme rainfall scenarios that include realistic estimates of rainfall duration, intensity, and space-time structure that can be used for probabilistic flood and landslide hazard and risk assessment at a wide range of scales.

The SST technique, as implemented in RainyDay, has two important features that distinguish it from IDF and design storm methods for describing the relationships between the intensity, duration, and structure of extreme rainfall. First, it leverages the detailed picture of rainfall space-time structure offered by ground-based radar or satellite-based sensors. This structure can play an important role in landslides and floods because the variability in the concentration and intermittency of extreme rainfall in space and time can lead to a complex and diverse spectrum of hazard response. This structure is difficult to measure using rain gages due to the high gage densities and sampling rates required, and so rain gage-based methods for analysis of rainfall-driven hazards, such as IDF relations and design storm methods, typically neglect this higher-order variability. The reader is directed to Wright et al. (2014b) for a deeper examination of this feature in the context of urban flood hazards.

The second important feature of RainyDay is that, because of the near-global coverage of satellite rainfall datasets, it is possible to generate realistic representations of extreme rainfall in remote or poorly-instrumented regions where rain gage or stream gage records are lacking. Such regions are common even in wealthy nations and are ubiquitous in developing countries, many of which are characterized by rapidly-growing exposure to rainfall-driven hazards due to urbanization and climate change. The authors are not aware of other approaches that offer the

ability to generate realistic rainfall inputs for probabilistic hazard modeling nearly anywhere on the globe with minimal computational effort.

Despite the advantages that SST and RainyDay offer over some other methods for assessing rainfall-driven hazards (e.g. design storms, discharge frequency analysis), a number of limitations and unanswered questions remain. Perhaps the biggest limitation to coupling SST with rainfall remote sensing, and to remote sensing applications more broadly, is the uncertain accuracy of the input rainfall data. Significant efforts have been made to better understand and minimize the errors in remote sensing estimates of rainfall, both from satellites (e.g. Petty and Krajewksi, 1996; Tian and Peters-Lidard, 2007; Tian et al., 2009) and from ground-based radar (e.g. Villarini and Krajewski, 2010). Such studies demonstrate that remote sensing estimates can vary significantly from reference observations in terms of rainfall intensity bias and differentiation between rainy and non-rainy areas, with important implications for hazard applications. In the case of satellite-based rainfall estimates, heterogeneities in the underlying land or water surfaces can be difficult to distinguish from variations in cloud and rainfall properties (e.g. Ferraro et al., 2013), while both ground-based radar and space-based sensors tend to suffer in mountainous areas due to dramatic variations in rainfall physical properties over short time and length scales. Furthermore, the spatial and temporal resolution of remote sensing estimates, particularly from satellites, can be too coarse for modeling at very small scales, especially in urban areas and fast-responding mountain or desert catchments where surface runoff generation from intense, short-duration rainfall on sub-hourly, sub-kilometer scales can be a key driver of hazards. The uncertainties associated with rainfall remote sensing data pose serious challenges for flood or landslide forecasting and monitoring, which require accurate rainfall estimates in real-time. These issues may be somewhat less critical in the SST framework or in long-term hazard assessment more generally, since the rainfall estimates need only have fidelity in the statistical sense. SST will be somewhat robust to random errors in rainfall data, as the underestimation of rainfall intensity from some storms in the storm catalog can be compensated by overestimation of rainfall intensity from others. In contrast, SST is not robust to systematic rainfall biases, as demonstrated in several examples in this paper. IMERG, NASA's newest satellite multi-sensor dataset, will feature improved accuracy and relatively high

resolution (0.1º, 30-minute), addressing some of these issues once the full retrospective dataset from 1998-present becomes available.

In the case of flood hazard modeling using SST, a practical upper limit on the size of the area of interest A can arise. As mentioned in Sections 2.1 and 3.3, the sizes of A and A' can be limited due to the challenges posed by transposition in the presence of complex terrain features. Furthermore, as A becomes larger, the rainfall duration t needed to properly model hazard response becomes longer. While RainyDay does not restrict the choice of t, practical limitations exist. In large watersheds, floods are usually the result of specific space-time arrangements of multiple distinct storm systems over the span of perhaps a week and up to months, often linked to persistent large-scale atmospheric disturbances. One could specify a long t (a month, for example) in RainyDay to "capture" all of these storm systems within a single storm catalog entry. Such long t, however, means there could only be relatively few entries in the storm catalog, given the limited record length of the input dataset. Such an approach would be constrained by the few space-time configurations of these storm systems that were actually observed, while other non-observed configurations are hypothetically possible. A tradeoff thus emerges as A (and thus t) increase relative to the area of the transposition domain A'. If A is a large fraction of A', then there is little opportunity to leverage the "space-for-time" substitution that is at the core of the SST approach. If the user instead decides to increase the size of A', he or she must ensure that this transposition is performed in a realistic manner. This effectively precludes modeling of regions that approach continental scales. The maximum scale at which SST can be feasibly used is an open question with no simple answer. It should be noted that IDF and design storm methods face similar and perhaps even more acute limitations in terms of an upper area limit, though for different reasons (e.g. conceptual and practical shortcomings of point-based temporal rainfall distributions and area reduction factors).

As mentioned in Section 4.3, a common critique of the methodology presented in this study is that the relatively short remote sensing records may not contain enough truly extreme rainfall events. Sensitivity to record length is not unique to SST; frequency estimates of rare hazards will be driven by the largest several events in the historical record, regardless of the chosen analysis technique. The results in Section 4.3 demonstrate that this concern may be somewhat

exaggerated in the case of SST since very extreme rainfall events that are considered rare from a local viewpoint can occur much more frequently when viewed regionally. Like more commonly-used regionalization techniques, SST helps to leverage this fact to improve hazard analysis. As the rainfall remote sensing record grows, the robustness of estimates produced by SST and RainyDay should increase as additional extreme storms are observed (and as their accuracy improves due to technological advances). Estimates of rainfall intensity will improve more per unit of time using SST than using point-based techniques due to SST's regional nature, while new patterns of rainfall space-time structure will add to the realism of SST-based flood and landslide hazard estimates since a broader spectrum of hazard outcomes will be possible. RainyDay makes such updating simple, while IDF databases and design storm methods are generally updated through slow and costly procedures (Y. Zhang, personal communication, May 14, 2015).

As highlighted in Section 4.2, SST and RainyDay have important features in the context of nonstationary hazards. Extreme rainfall scenarios from RainyDay are generally based on more recent observations than existing rain gage or stream gage-based frequency analyses such as Atlas 14 IDF relations, which contain older records that may not be representative of the current state of the climate. In this respect, hazard analysis based on RainyDay can be understood as a relatively current "snapshot" based on recent climate. The performance of RainyDay is very dependent on major storms having occurred one or more times within the transposition domain, however, meaning that spatial transposition is not a perfect remedy for short data records. Furthermore, if the rainfall remote sensing record deviates significantly from the true long-term properties of extreme rainfall over the region of interest due to random chance, decadal-scale climate variability, or systematic measurement bias, then caution must be taken when using RainyDay. It can be challenging in practice to diagnose such nonstationarities and biases due to a lack of long-term independent observational data, particularly in remote or underdeveloped regions. Meanwhile, as discussed in Wright et al. (2014b), combining SST (or other rainfall-based approaches, e.g. Cunha et al., 2011) with a distributed hazard model allows the analyst to incorporate changes in land use and land cover into hazard estimates.

Acknowledgments

This work was made possible through the fellowship support of the NASA Postdoctoral Program, administered by Oak Ridge Associated Universities, Oak Ridge, Tennessee. We also acknowledge the support of the University of Wisconsin and the Iowa Flood Center and the University of Iowa. We would like to acknowledge the help of Soni Yatheendradas and Dalia Kirschbaum of NASA Goddard Space Flight Center for provision of the PUSH rainfall error modeling code, which ultimately was not included as part of this manuscript nor in the current version of RainyDay, but which helped to illuminate the challenges of incorporating stochastic rainfall error models into the SST framework. We would also like to thank Scott Small, Chi Chi Choi, and Tibebu Ayalew at the University of Iowa for their support in configuring and troubleshooting the IFC Model. Computing resources supporting the hydrologic modeling were provided by the NASA High-End Computing Program through the NASA Center for Climate Simulation at Goddard Space Flight Center. We would also like to thank the editor and the two anonymous reviewers whose constructive criticisms contributed greatly to the study.

References

Alexander, G.N., 1963. Using the probability of storm transposition for estimating the frequency of rare floods. J. Hydrol. 1, 46–57.

Bonnin, G.M., Martin, D., Lin, B., Parzybok, T., Riley, D., 2004. NOAA Atlas 14: Precipitation-Frequency Atlas of the United States.

Brenning, A., 2005. Spatial prediction models for landslide hazards: review, comparison and evaluation. Nat. Hazards Earth Syst. Sci. 5, 853–862.

Burnash, R.J.C., 1995. The NWS river forecast system: Catchment modeling. Comput. Model. Watershed Hydrol. 311–366.

Chen, M., Shi, W., Xie, P., Silva, V.B.S., Kousky, V.E., Wayne Higgins, R., Janowiak, J.E., 2008. Assessing objective techniques for gauge-based analyses of global daily precipitation. J. Geophys. Res. Atmos. 113, n/a–n/a.

Cheng, L., AghaKouchak, A., 2014. Nonstationary precipitation Intensity-Duration-Frequency curves for infrastructure design in a changing climate. Sci. Rep. 4, 7093.

Ciach, G.J., Morrissey, M.L., Krajewski, W.F., 2000. Conditional Bias in Radar Rainfall

1168 Estimation. J. Appl. Meteor. 39, 1941–1946.

1169 Crum, T.D., Alberty, R.L., 1993. The WSR-88D and the WSR-88D Operational Support
1170 Facility. Bull. Am. Meteorol. Soc. 74, 1669–1687.

1171 Cunha, L.K., Krajewski, W.F., Mantilla, R., Cunha, L., 2011. A framework for flood risk
1172 assessment under nonstationary conditions or in the absence of historical data. J. Flood Risk
1173 Manag. 4, 3–22.

1174 Cunha, L.K., Mandapaka, P. V, Krajewski, W.F., Mantilla, R., Bradley, A.A., 2012. Impact of
1175 radar-rainfall error structure on estimated flood magnitude across scales: An investigation
1176 based on a parsimonious distributed hydrological model. Water Resour. Res. 48.

1177 Cunnane, C., 1978. Unbiased plotting positions — A review. J. Hydrol. 37, 205–222.

1178 Demir, I., Krajewski, W.F., 2013. Towards an integrated Flood Information System: Centralized
1179 data access, analysis, and visualization. Environ. Model. Softw. 50, 77–84.

1180 Eash, D.A., Barnes, K.K., Veilleux, A.G., 2013. Methods for estimating annual exceedance-
1181 probability discharges for streams in Iowa, based on data through water year 2010: U.S.
1182 Geological Survey Scientific Investigations Report 2013-5086.

1183 Ebert, E.E., Janowiak, J.E., Kidd, C., 2007. Comparison of near-real-time precipitation estimates
1184 from satellite observations and numerical models. Bull. Am. Meteorol. Soc. 88, 47–64.

1185 England, J.F., Julien, P.Y., Velleux, M.L., 2014. Physically-based extreme flood frequency with
1186 stochastic storm transposition and paleoflood data on large watersheds. J. Hydrol. 510, 228–
1187 245.

1188 Ferraro, R.R., Peters-Lidard, C.D., Hernandez, C., Turk, F.J., Aires, F., Prigent, C., Lin, X.,
1189 Boukabara, S.-A., Furuzawa, F.A., Gopalan, K., Harrison, K.W., Karbou, F., Li, L., Liu, C.,
1190 Masunaga, H., Moy, L., Ringerud, S., Skofronick-Jackson, G.M., Tian, Y., Wang, N.-Y.,
1191 2013. An Evaluation of Microwave Land Surface Emissivities Over the Continental United
1192 States to Benefit GPM-Era Precipitation Algorithms. IEEE Trans. Geosci. Remote Sens. 51,
1193 378–398.

1194 Fontaine, T.A., Potter, K.W., 1989. Estimating Probabilities of Extreme Rainfalls. J. Hydraul.
1195 Eng. 115, 1562–1575.

1196 Foufoula-Georgiou, E., 1989. A probabilistic storm transposition approach for estimating
1197 exceedance probabilities of extreme precipitation depths. Water Resour. Res. 25, 799–815.

1198 Franchini, M., Helmlinger, K.R., Foufoula-Georgiou, E., Todini, E., 1996. Stochastic storm
1199 transposition coupled with rainfall-runoff modeling for estimation of exceedance
1200 probabilities of design floods. J. Hydrol. 175, 511–532.

1201 Gupta, V.K., 1972. Transposition of Storms for Estimating Flood Probability Distributions.
1202 Colorado State University.

1203 Habib, E., Henschke, A., Adler, R.F., 2009. Evaluation of TMPA satellite-based research and
1204 real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana,
1205 USA. Atmos. Res. 94, 373–388.

1206 Horritt, M.S., Bates, P.D., 2002. Evaluation of 1D and 2D numerical models for predicting river
1207 flood inundation. J. Hydrol. 268, 87–99.

1208 Huffman, G.J., Adler, R.F., Bolvin, D.T., Nelkin, E.J., 2010. The TRMM Multi-satellite
1209 Precipitation Analysis (TMPA). In: Hossain, F., Gebremichael, M. (Eds.), Satellite Rainfall
1210 Applications for Surface Hydrology. Springer Verlag, pp. 3–22.

1211 Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R.J., Xie, P., 2014. Algorithm
1212 Theoretical Basis Document (ATBD) Version 4- NASA Global Precipitation Measurement
1213 (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG), PMM Website.

1214 Interagency Advisory Committee on Water Data (IACWD), 1982. Guidelines for determining
1215 flood flow frequency, Bulletin 17B. Reston, VA.

1216 Jones, E., Oliphant, T., Peterson, P., Walt, S. van der, Colbert, S.C., Varoquaux, G., 2011. SciPy:
1217 Open source scientific tools for Python. Comput. Sci. Eng. 13.

1218 Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: A method that produces
1219 global precipitation estimates from passive microwave and infrared data at high spatial and
1220 temporal resolution. J. Hydrometeorol. 5, 487–503.

1221 Lin, Y., Mitchell, K.E., 2005. The NCEP Stage II/IV hourly precipitation analyses: development
1222 and applications. In: Preprints, 19th Conf. on Hydrology, American Meteorological Society,
1223 San Diego, CA, 9-13 January 2005, Paper 1.2. pp. 2–5.

1224 Mallakpour, I., Villarini, G., 2015. The changing nature of flooding across the central United
1225 States. Nat. Clim. Chang. 5, 250–254.

1226 Mantilla, R., Gupta, V.K., 2005. A GIS numerical framework to study the process basis of
1227 scaling statistics in river networks. Geosci. Remote Sens. Lett. IEEE 2, 404–408.

1228 McCuen, R.H., 1998. Hydrologic analysis and design, Mcgraw- Hill.

1229 Mehran, A., AghaKouchak, A., 2014. Capabilities of satellite precipitation datasets to estimate
1230 heavy precipitation rates at different temporal accumulations. Hydrol. Process. 28, 2262–
1231 2270.

1232 Moser, B.A., Gallus Jr., W.A., Mantilla, R., 2015. An initial assessment of radar data
1233 assimilation on warm season rainfall forecasts for use in hydrologic models. Weather
1234 Forecast. 30, 1491–1520.

1235 Nikolopoulos, E.I., Anagnostou, E.N., Borga, M., 2013. Using High-Resolution Satellite Rainfall
1236 Products to Simulate a Major Flash Flood Event in Northern Italy. J. Hydrometeorol. 14,
1237 171–185.

1238 Petersen-Øverleir, A., 2004. Accounting for heteroscedasticity in rating curve estimates. J.
1239 Hydrol. 292, 173–181.

1240 Petersen-Øverleir, A., Reitan, T., 2009. Accounting for rating curve imprecision in flood
1241 frequency analysis using likelihood-based methods. J. Hydrol. 366, 89–100.

1242 Petty, G.W., Krajewksi, W.F., 1996. Satellite estimation of precipitation over land. Hydrol. Sci.
1243 J. 41, 433–451.

1244 Potter, K.W., Walker, J.F., 1985. An Empirical Study of Flood Measurement Error. Water
1245 Resour. Res. 21, 403–406.

1246 Preisig, M., Zimmermann, T., 2010. Two-phase free-surface fluid dynamics on moving domains.
1247 J. Comput. Phys. 229, 2740–2758.

1248 Rogger, M., Kohl, B., Pirkl, H., Viglione, A., Komma, J., Kirnbauer, R., Merz, R., Blöschl, G.,
1249 2012. Runoff models and flood frequency statistics for design flood estimation in Austria –
1250 Do they tell a consistent story? J. Hydrol. 456-457, 30–43.

1251 Sen, P.K., 1968. Estimates of the regression coefficient based on Kendall's tau. J. Am. Stat.

Assoc. 63, 1379–1389.

Shige, S., Kida, S., Ashiwake, H., Kubota, T., Aonashi, K., 2013. Improvement of TMI rain retrievals in mountainous areas. J. Appl. Meteorol. Climatol. 52, 242–254.

Small, S.J., Jay, L.O., Mantilla, R., Curtu, R., Cunha, L.K., Fonley, M., Krajewski, W.F., 2013. An asynchronous solver for systems of ODEs linked by a directed tree structure. Adv. Water Resour. 53, 23–32.

Smith, J.A., Baeck, M.L., Villarini, G., Wright, D.B., Krajewski, W., 2013. Extreme Flood Response: The June 2008 Flooding in Iowa. J. Hydrometeorol. 14, 1810–1825.

Smith, J.A., Villarini, G., Baeck, M.L., 2011. Mixture Distributions and the Hydroclimatology of Extreme Rainfall and Flooding in the Eastern United States. J. Hydrometeorol. 12, 294–309.

Smith, M.B., Seo, D.-J., Koren, V.I., Reed, S.M., Zhang, Z., Duan, Q., Moreda, F., Cong, S., 2004. The distributed model intercomparison project (DMIP): motivation and experiment design. J. Hydrol. 298, 4–26.

Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H. V., Imam, B., Braithwaite, D., 2000. Evaluation of PERSIANN System Satellite-Based Estimates of Tropical Rainfall. Bull. Am. Meteorol. Soc. 81, 2035–2046.

Stampoulis, D., Anagnostou, E.N., Nikolopoulos, E.I., 2013. Assessment of High-Resolution Satellite-Based Rainfall Estimates over the Mediterranean during Heavy Precipitation Events. J. Hydrometeor 14, 1500–1514.

Tian, Y., Peters-Lidard, C.D., 2007. Systematic anomalies over inland water bodies in satellite-based precipitation estimates. Geophys. Res. Lett. 34, L14403.

Tian, Y., Peters-Lidard, C.D., Choudhury, B.J., Garcia, M., 2007. Multitemporal Analysis of TRMM-Based Satellite Precipitation Products for Land Data Assimilation Applications. J. Hydrometeorol. 8, 1165–1183.

Tian, Y., Peters-Lidard, C.D., Eylander, J.B., Joyce, R.J., Huffman, G.J., Adler, R.F., Hsu, K., Turk, F.J., Garcia, M., Zeng, J., 2009. Component analysis of errors in satellite-based precipitation estimates. J. Geophys. Res. 114, D24101.

Troutman, B.M., Karlinger, M.R., 2003. Regional flood probabilities. Water Resour. Res. 39, 1095.

U.S. Weather Bureau, 1958. Rainfall intensity-frequency regime, Part 2-Southeastern United States, Technical Paper No. 29.

Villarini, G., Krajewski, W.F., 2010. Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall. Surv. Geophys. 31, 107–129.

Villarini, G., Smith, J.A., Baeck, M.L., Vitolo, R., Stephenson, D.B., Krajewski, W.F., 2011. On the frequency of heavy rainfall for the Midwest of the United States. J. Hydrol. 400, 103–120.

Villarini, G., Smith, J.A., Vitolo, R., Stephenson, D.B., 2013. On the temporal clustering of US floods and its relationship to climate teleconnection patterns. Int. J. Climatol. 33, 629–640.

Villarini, G., Strong, A., 2014. Roles of climate and agricultural practices in discharge changes in an agricultural watershed in Iowa. Agric. Ecosyst. Environ. 188, 204–211.

Walshaw, D., 2013. Generalized Extreme Value DistributionBased in part on the article "Generalized extreme value distribution" by Jan Beirlant and Gunther Matthys, which appeared in the Encyclopedia of Environmetrics. In: Encyclopedia of Environmetrics. John Wiley & Sons, Ltd, Chichester, UK.

Walt, S. van der, Colbert, S.C., Varoquaux, G., 2011. The NumPy Array: A Structure for Efficient Numerical Computation. Comput. Sci. Eng. 13.

Wilson, L.L., Foufoula-Georgiou, E., 1990. Regional Rainfall Frequency Analysis via Stochastic Storm Transposition. J. Hydraul. Eng. 116, 859–880.

Wright, D.B., Smith, J.A., Baeck, M.L., 2014a. Critical Examination of Area Reduction Factors. J. Hydrol. Eng. 19, 769–776.

Wright, D.B., Smith, J.A., Baeck, M.L., 2014b. Flood frequency analysis using radar rainfall fields and stochastic storm transposition. Water Resour. Res. 50, 1592–1615.

Wright, D.B., Smith, J.A., Villarini, G., Baeck, M.L., 2013. Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. J. Hydrol. 488, 150–165.

Wright, D.B., Smith, J.A., Villarini, G., Baeck, M.L., 2014c. Long-Term High-Resolution Radar

Rainfall Fields for Urban Hydrology. JAWRA J. Am. Water Resour. Assoc. 50, 713–734.

Xie, P., Yatagai, A., Chen, M., Hayasaka, T., Fukushima, Y., Liu, C., Yang, S., 2007. A Gauge-Based Analysis of Daily Precipitation over East Asia. J. Hydrometeorol. 8, 607.

Yarnell, D.L., 1935. Rainfall Intensity-Frequency Data. Washington, D. C.